REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE Feb 95	3. REPORT TYPE AND DATES COVERED Final 1 Jan 94 – 31 Dec 94

4. TITLE AND SUBTITLE

1994 Sanibel Symposium on Quantum Physics

5. FUNDING NUMBERS

DAAH04-94-G-0015

6. AUTHOR(S)

Rodney J. Bartlett

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Florida
Gainesville, FL 32611

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO 32557.1-PH-CF

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

DTIC
ELECTE
MAR 0 9 1995
B
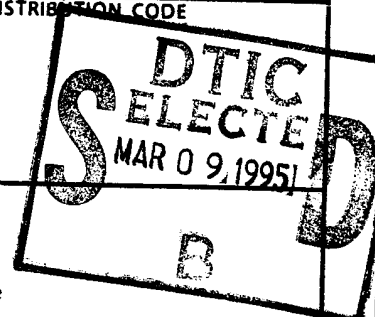
13. ABSTRACT (Maximum 200 words)

The 34th Annual Sanibel Symposium, organized by the faculty and staff of the Quantum Theory Project of the University of Florida, was held on February 12–19, 1994, at the Marriott, Sawgrass Resort, Ponte Vedra Beach, Florida. Over 300 participants gathered for 8 days of lectures and informal discussions.

The format of the symposium adopted for the past few years was followed again this year with a compact 8-day schedule with an integrated program of quantum biology, quantum chemistry, and condensed matter physics. The topics of the sessions covered by these proceedings include Quantum Chemistry of Biological Molecules, Spectroscopic Signatures of Biological Molecules, Protein Folding, and Photosynthesis.

"Original contains color plates: All DTIC reproductions will be in black and white"

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

International Journal of quantum chemistry

QUANTUM BIOLOGY

SYMPOSIUM NO. 21, 1994

Proceedings of the
International Symposium on the
Application of Fundamental Theory to

Problems of Biology

and Pharmacology

Proceedings of the
International Symposium on the
Application of Fundamental
Theory to

Problems of Biology
and Pharmacology

Held at Ponte Vedra Beach, Florida, February 12–19, 1994

Editor-in-Chief: Per-Olov Löwdin
Special Editors: N. Yngve Öhrn, John R. Sabin, and
Michael C. Zerner

Contents

iv   CONTENTS

Introduction

The 34th Annual Sanibel Symposium, organized by the faculty and staff of the Quantum Theory Project of the University of Florida, was held on February 12–19, 1994, at the Marriott, Sawgrass Resort, Ponte Vedra Beach, Florida. Over 300 participants gathered for 8 days of lectures and informal discussions.

The format of the symposium adopted for the past few years was followed again this year with a compact 8-day schedule with an integrated program of quantum biology, quantum chemistry, and condensed matter physics. The topics of the sessions covered by these proceedings include Quantum Chemistry of Biological Molecules, Spectroscopic Signatures of Biological Molecules, Protein Folding, and Photosynthesis.

The articles have been subjected to the ordinary refereeing procedures of the *International Journal of Quantum Chemistry*. The articles presented in the sessions on quantum chemistry, condensed matter physics, and associated poster sessions are published in a separate volume of the *International Journal of Quantum Chemistry*.

The organizers acknowledge the following sponsors for their support of the 1994 Sanibel Symposium:

- The Office of Naval Research through Grant NO0014-93-1-0343
  "This work relates to Department of Navy Grant NO0014-93-1-0343 issued by the Office of Naval Research. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein."
- U.S. Army Research Office (Physics)/CRDEC and U.S. Army Edgewood RD&E Center through Grant DAAH04-94-G-0015
  "The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation."
- U.S. Department of Energy through Grant DE-FG05-94ER61785
- International Science Foundation
- CAChe
- IBM
- International Society of Quantum Biology and Pharmacology
- Silicon Graphics
- Sun Microsystems
- The University of Florida.

Very special thanks go to the staff of the Quantum Theory Project of the University of Florida for handling the numerous administrative, clerical, and practical details.

The organizers are proud to recognize the contributions of Mrs. Judy Parker, Ms. Leann Golemo, Mrs. Karen Yanke, Ms. Sandra Weakland, Mr. Sullivan Beck, Dr. Agustín Diz, and Dr. Erik Deumens. All the graduate students of the Quantum Theory Project who served as "gofers" are gratefully recognized for their contributions to the 1994 Sanibel Symposium.

N. Y. ÖHRN
J. R. SABIN
M. C. ZERNER

Postulates for Protein (Hydrophobic) Folding and Function

DAN W. URRY

Laboratory of Molecular Biophysics, School of Medicine, The University of Alabama at Birmingham, VH 300, Birmingham, Alabama 35294-0019

Abstract

The previously demonstrated capacity to utilize the hydrophobic folding and assembly transition in designed model proteins to perform diverse energy conversions is formalized in terms of three postulates with an associated 15 corollaries. Each corollary defines one of the 15 pairwise free-energy transductions involving the six intensive variables: mechanical force, temperature, pressure, chemical potential, electrochemical potential, and electromagnetic radiation. The first postulate directly involves the input of thermal energy to raise the temperature from below to above that temperature required to drive hydrophobic folding and assembly, with the resultant capacity to perform useful mechanical work. The second postulate considers the energy inputs that can lower or raise the temperature range over which the hydrophobic folding transition occurs; these energy inputs can thereby perform mechanical work. The third postulate treats the energy conversions not involving mechanical force, whereby a pair of functional groups becomes coupled by each being a part of the same hydrophobic association process with each functional group being able individually to drive the hydrophobic folding transition and thereby to change the state of the second functional group in an energy-conversion process. It is then shown how a model protein can be so designed as to function in a second-order process treated by Postulate III and how the efficiency of energy conversion can be enhanced. Finally, these energy-conversion studies using the hydrophobic folding and assembly transition in model proteins are related to a theoretical model for cooperative hydrophobic folding and to experimental studies suggesting a central role for hydrophobic folding in the general problem of protein folding. © 1994 John Wiley & Sons, Inc.

Protein Folding and Function

It is an aphorism that understanding protein folding is essential to understanding protein function. Although, commonly, protein function has been considered in terms of active sites and the breaking and forming of bonds therein, much of protein function in living organisms can be viewed in terms of the conversion of energy from an available form to a more directly useful form. Indeed, this is a key step in the evolution of living organisms, and it can be achieved by controlling protein folding. Most obviously, light and chemical stores such as starches and fats are recognized as sources of energy for living organisms and, of course, there is the universal immediate chemical energy source, adenosine triphosphate (ATP).

A construct that can convert energy from one form to another can be called a machine. The molecular constructs, proteins, could therefore be reasonably called molecular machines. Those particular molecular machines that can convert energy into useful mechanical motion could be called molecular engines. The actin and

myosin molecular construct of muscle could be considered a molecular engine as this molecular construct gives rise to useful mechanical motion on using ATP as an energy source. But there are many other energies at play in living organisms. There are some six forms of free energy involved in the functioning of proteins in living organisms, the intensive variables of which are mechanical force, temperature, pressure, chemical potential, electrochemical potential, and electromagnetic radiation, and proteins are the molecular machines that perform these energy conversions. Interestingly, as will be presented below, hydrophobic folding of proteins, seen in terms of certain model proteins, appears to be an element of folding that can perform the 15 pairwise free-energy transductions possible given the above six intensive variables of the free energy [1].

Postulates for Protein-Catalyzed Energy Conversion Using Hydrophobic Folding

The following analysis has an underlying hypothesis that all the energy conversion functions of proteins can be achieved by controlling hydrophobic folding and assembly. This hypothesis has as its origin in the literature a paper presented at and occurring in the *Proceedings of the 1987 Sanibel Symposium* [2]. This hypothesis can now be stated in terms of three postulates and an associated 15 corollaries. These are listed in Table I.

Model Proteins as First-Order Molecular Machines of the T_t *Type: Molecular Engines*

Hydrophobic folding and assembly of amphiphilic polymers like proteins are such that at sufficiently low temperature in water the hydrophobic moieties are surrounded by a water of hydrophobic hydration. Thermodynamically, this water of hydrophobic hydration is of low entropy and is exothermic in its formation [3–10], and, structurally, it is characterized by water molecules arranged at the corners of pentagons with some 12 such pentagons able to enclose a hydrophobic entity such as a methane molecule [11,12].

Depending on the particular composition of a protein-based polymer construct containing both hydrophobic and polar components, as the temperature is raised through a discrete transition temperature range, the pentagonally arranged waters of hydrophobic hydration destructure in an endothermic reaction as the protein part of the system folds and associates using the now available hydrophobic intramolecular and intermolecular contacts [9,10,12–16]. This is referred to as an inverse temperature transition because the protein part of the system increases order with increase in temperature through the transition temperature range, the onset temperature for which is designated as T_t [1,17].

Postulate I. Using elastic model proteins, also referred to as elastic protein-based polymers of the composition $poly[f_V(VPGVG), f_X(VPGXG)]$, where f_X and f_V are mole fractions, with $f_V + f_X = 1$, it has been possible to observe this inverse temperature transition process in wholly synthetic systems by γ-irradiation cross-linking to form macroscopic elastic matrices. At temperatures lower than T_t, the elastic matrix is swollen; on raising the temperature above T_t through the temper-

TABLE I. Postulates for energy conversion by means of hydrophobic folding and assembly (inverse temperature transitions).

Postulate I: The input of thermal energy to a protein capable of *hydrophobic folding and assembly* on raising the temperature from below to above the temperature, T_t, of an *inverse temperature transition* can result in motion and the performance of mechanical work

    Corollary: Thermomechanical transduction

Postulate II: Any energy input that changes the temperature, T_t, at which an inverse temperature transition occurs can be used to produce motion and perform mechanical work

    Corollary 1: Chemomechanical transduction
    Corollary 2: Electromechanical transduction
    Corollary 3: Baromechanical transduction
    Corollary 4: Photomechanical transduction

Postulate III: Different energy inputs, each of which can individually drive hydrophobic folding to produce motion and perform mechanical work, can be converted one into the other (transduced) by means of the inverse temperature transition with the correctly designed coupling and T_t value

    Corollary 1: Electrochemical transduction
    Corollary 2: Electrothermal transduction
    Corollary 3: Baroelectrical transduction
    Corollary 4: Photovoltaic transduction
    Corollary 5: Thermochemical transduction
    Corollary 6: Photothermal transduction
    Corollary 7: Barothermal transduction
    Corollary 8: Barochemical transduction
    Corollary 9: Photobaric transduction
    Corollary 10: Photochemical transduction

ature range of the inverse temperature transition, the elastic matrix contracts to less than one-half its swollen dimension [18,19]. When a weight is hung on the swollen elastic matrix and the temperature is raised above T_t, the matrix contracts and lifts the weight. Thus, this construct that hydrophobically folds and assembles on raising the temperature is a molecular machine capable of performing useful mechanical motion; it is a molecular engine. Thermal energy has been converted into mechanical work. The process is thermomechanical transduction.

A T_t-based Hydrophobicity Scale. By systematically varying X in poly-[f_V(VPGVG),f_X(VPGXG)], a unique hydrophobicity scale has been developed that is based directly on the hydrophobic folding process of interest. The value of T_t for poly(VPGVG), i.e., for $f_V = 1$, is 25°C. When X = Ile(I) and $f_X = 1$ as in poly(VPGIG), the addition of the one CH_2 moiety per pentamer lowers the value of T_t some 15°C. When X = Ala(A) as in poly(VPGAG), the loss of two CH_2 moieties per pentamer raises the value of T_t some 30°C. A plot of f_X vs. T_t is given in Figure 1 for poly[f_V(VPGVG),f_X(VPGXG)] in phosphate buffered saline and

Temperature of Inverse Temperature Transition, T_t
for poly[f_V(VPGVG),f_X(VPGXG)]
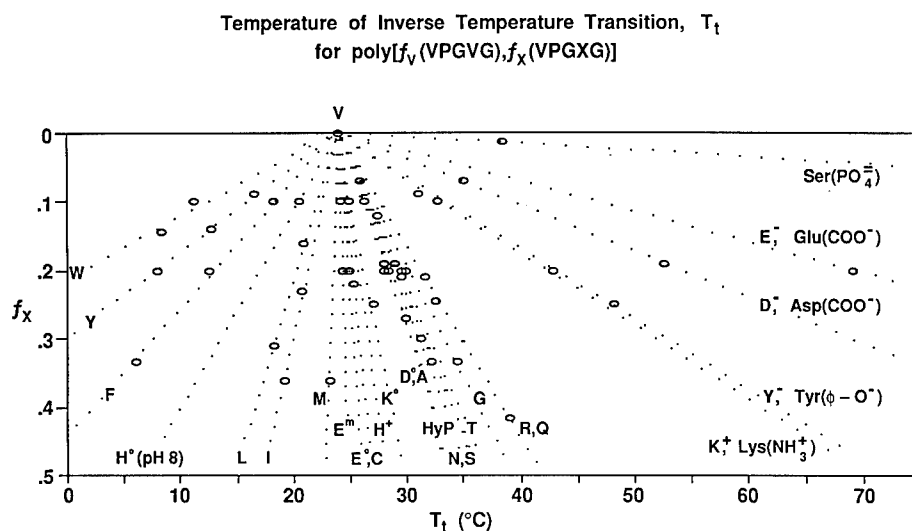
Figure 1.   Plot of the mole fraction, f_X, of pentamers containing the guest residue X vs.
the temperature, T_t, for the onset of the hydrophobic folding and assembly transition for
model proteins of the structure poly[f_V(VPGVG),f_X(VPGXG)], where f_V and f_X are the
mole fractions with $f_V + f_X = 1$ and where X is any naturally occurring amino acid residue
or a chemical modification thereof. Reproduced with permission from [20].

the values of T_t extrapolated to $f_X = 1$ are given in Table II. The more hydrophobic
a residue is the lower is the value of T_t and the less hydrophobic the higher the
value of T_t [1,20].

Postulate II and the ΔT_t Mechanism of Energy Conversion. Note in Table II and
Figure 1 that the values of T_t are very different for the Glu(E) residue when the
carboxyl side chain is protonated as COOH (where $T_t = 30°$C) than when the side
chain is ionized as for COO$^-$ (where $T_t = 250°$C). This means that at 37°C a cross-
linked elastic matrix of poly[0.8(VPGVG),0.2(VPGEG)] will be contracted at
pH 3 and swollen at neutral pH. Thus, it is possible to attach a weight to the swollen
elastic matrix at 37°C and pH 7 and to lower the pH to 3 and cause contraction
and the lifting of the weight [21]. This involves the chemical energy input of raising
the proton concentration (increasing the proton chemical potential) with the re-
sulting performance of useful mechanical work. This is chemomechanical trans-
duction (Corollary 1 of Table I).

It may also be noted in Table II, for the *N*-methyl nicotinamide moiety attached
to a lysine, Lys(K), side chain by amide linkage, that the value of T_t is 120°C
when oxidized but −130°C when reduced. Thus, it is possible in the cross-linked
protein-based polymer of poly[0.73(VPGVG),0.27(VPGK{NMeN}G)] at 37°C
to attach a weight to the swollen (oxidized) elastic matrix and to reduce the nico-
tinamide to drive contraction and perform useful mechanical work [1]. This can
be the input of electrical energy with the result of useful mechanical work; it is
electromechanical transduction (Corollary 2 of Table I).

TABLE II. T_t-based hydrophobicity scale for proteins T_t = temperature of inverse temperature transition for poly[f_v(VPGVG), f_x(VPGXG)].[a]

Residue X		T_t, linearly extrapolated to $f_x = 1$ (°C)	Correlation coefficient
Lys(NMeN, reduced)[b]		−130	1.000
Trp	(W)	−90	.993
Tyr	(Y)	−55	.999
Phe	(F)	−30	.999
His (pH 8)	(H⁰)	−10	1.000
Pro	(P)[c]	(−8)	Calculated
Leu	(L)	5	.999
Ile	(I)	10	.999
Met	(M)	20	.996
Val	(V)	24	Reference
Glu(COOCH₃)	(Eᵐ)	25	1.000
Glu(COOH)	(E⁰)	30	1.000
Cys	(C)	30	1.000
His (pH 4)	(H⁺)	30	1.000
Lys(NH₂)	(K⁰)	35	.936
Pro	(P)[d]	40	.950
Asp(COOH)	(D⁰)	45	.994
Ala	(A)	45	.997
HyP		50	.998
Asn	(N)	50	.997
Ser	(S)	50	.997
Thr	(T)	50	.999
Gly	(G)	55	.999
Arg	(R)	60	1.000
Gln	(Q)	60	.999
Lys(NH₃⁺)	(K⁺)	120	.999
Tyr(φ-O⁻)	(Y⁻)	120	.996
Lys(NMeN, oxidized)[b]		120	1.000
Asp(COO⁻)	(D⁻)	170	.999
Glu(COO⁻)	(E⁻)	250	1.000
Ser(PO₄²⁻)		1000	1.000

[a] Adapted with permission from [20].

[b] NMeN is for *N*-methyl nicotinamide pendant on a lysyl side chain, i.e., *N*-methyl nicotinate attached by amide linkage to the ε-NH₂ of Lys and the reduced state is *N*-methyl-1,6-dihydronicotinamide.

[c] The calculated T_t value for Pro comes from poly(VPGVG) when the experimental values of Val and Gly are used. This hydrophobicity value of −8°C is unique to the β-spiral structure where there is hydrophobic contact between the Val¹γCH₃ and Pro²βCH₂ moieties.

[d] The experimental value determined from poly[f_x(VPGVG), f_p(PPGVG)].

It has also been found that the application of pressure particularly when X is an aromatic residue such as Phe(F), Tyr(Y), or Trp(W) will raise the value of T_t. Accordingly, in the appropriate elastic matrix and at a temperature just above T_t, the application of pressure will cause swelling with the lowering of an attached

weight and the release of pressure will cause a contraction with the lifting of the weight [22]. This is baro-mechanical transduction (Corollary 3 of Table I).

Most recently, it has been demonstrated that when azobenzene is attached to a Glu(E) side chain the appropriate wavelength of light will reversibly increase the value of T_t due to a light-driven *trans* → *cis* geometrical isomerism, where the *cis* state is less hydrophobic than is the *trans* state [23]. This provides a basis for photomechanical transduction (Corollary 4 of Table I). A similar effect has been found when cinnamic acid is the chromophore attached to a Lys(K) residue (Heimbach et al., unpublished data).

The phenomenological basis for Postulate II is that a properly designed protein-based polymer construct will change its value of T_t as the result of a particular energy input. This is called the ΔT_t mechanism of free-energy transduction because, e.g., the lowering of T_t can cause contraction and the performance of useful motion [1].

Model Proteins as Second-Order Molecular Machines of the T_t *Type*

In the above five conversions of free energy, i.e., of thermo-, chemo-, electro-, baro-, and photomechanical transduction, it was the hydrophobic folding process that directly performed the useful mechanical work. These molecular engines were therefore termed first-order molecular machines of the T_t type, where T_t type indicated the inverse temperature transition of hydrophobic folding and assembly on raising the temperature from below to above the transition temperature range. In terms of Figure 2, these are the five pairwise energy-conversion arrows that have one end at the mechanical force apex.

Postulate III. Now it is also possible to use the hydrophobic folding and assembly process to perform (catalyze) energy conversions not macroscopically involving the intensive variable of mechanical force. These energy conversions involve the other 10 pairwise energy conversions of Figure 2 not involving the mechanical force apex, and they are the 10 Corollaries of Postulate III, which, itself, states that "different energy inputs, each of which can individually drive hydrophobic folding to produce motion and perform mechanical work, can be converted one into the other (transduced) by means of the inverse temperature transition with the correctly designed coupling of functional moieties and T_t value." Energy conversions of this T_t type can involve the proper balancing of mean hydrophobicity of a potential hydrophobic domain with a resident pair of different functional groups in their more polar states such that T_t is just above the desired operating temperature, and it can involve designs for enhanced efficiency for energy conversion. The design process to achieve this balancing and enhanced efficiency is called *poising,* as will be discussed below.

Poising and Enhanced Efficiency of Energy Conversion

To poise is to draw up into readiness. With regard to the design of second-order elastic molecular machines of the T_t type, it is to so design an amphiphilic polymer such that a change in one form of free energy can trigger a change in a second form

Demonstrated and Putative Energy Conversions Using Molecular Machines of the T_t-type

mechanical
$(f\Delta L, L\Delta f)$

photo-mechanical

baro-mechanical

electromagnetic radiation (hv)

photo-baric

pressure
$(P\Delta V, V\Delta P)$

electro-mechanical

photo-chemical

piezo-electric effect

thermo-mechanical

chemo-mechanical

photo-voltaic

redox

baro-chemical

photo-thermal

electrical
$(zFE_0'\Delta n', n'zF\Delta E_0')$

electro-chemical

baro-thermal

chemical
$\mu\Delta n, n\Delta\mu$

electro-thermal

thermo-chemical

thermal
$(T\Delta S, S\Delta T)$

observed energy conversions

feasibility demonstrated by ΔT_t

energy conversions indirectly demonstrated

hypothetical, yet to be demonstrated

Figure 2.   Representation of the demonstrated and putative energy conversions using the hydrophobic folding and assembly transition. Adapted with permission from [1].

of free energy. For example, the reduction of an oxidized prosthetic group involving the free-energy change, $zFE_0'\Delta n'$, noted in the electrical apex of Figure 2 can cause an increase in the pK_a from below to above the solution pH of a second functional moiety, e.g., a $COOH/COO^-$ chemical couple, resulting in protonation and a decrease in the chemical potential of protons in solution. Or, conversely, it could be the protonation of a carboxylate that would change the redox potential of a prosthetic group. It should be emphasized that the protonatable and reducible functional groups are separate chemical entities that could be far removed in terms of covalent structure but which have in common that they are part of the same hydrophobic folding domain.

Balancing of Hydrophobicity and a Pair of Functional Groups in Their More Polar States

This balancing of hydrophobicity and a pair of functional groups in their more polar states can be depicted in terms of temperature profiles for folding and/or

aggregation as shown in Figure 3. The composition of the elastic protein-based polymer is designed such that, with a Glu(E) or Asp(D) residue with the carboxylate side chain and with a N-methyl nicotinamide attached to a Lys(K) side chain in its oxidized state, the hydrophobic folding transition begins just above the operating temperature, T_0. If the nicotinamide is reduced as in Figure 3(A), the temperature for the hydrophobic folding transition will be lowered to well below T_0, causing an increase in the pK_a of the Glu or Asp residue above that of the solution pH such that folding will occur with the protonation of the carboxylate, i.e., reduction results in a ΔpK_a and the uptake of a proton. If the carboxylate is protonated as in Figure 3(B), the drive for hydrophobic folding causes a change in the redox potential, a $\Delta E_0'$, such that under the appropriate condition there can be the reduction of the nicotinamide, i.e., protonation can result in the uptake of an electron. As seen Figure 3, this is again the result of the ΔT_t mechanism.

Nonlinear Hydrophobic-induced pKa Shifts

As the model protein becomes more hydrophobic, the change in pK_a resulting from a given change in hydrophobicity, ΔHPB, becomes larger. Thus, in poly[f_V(IPGVG),f_X(IPGXG)], where X can be Glu(E) or Asp(D), the change in hydrophobicity represented by replacement, e.g., of one Glu by a Val residue per 30 residues, has a much greater effect on the pK_a of the Glu residue when the change is from 2 Glu residues to 1 Glu residue per 30 residues than from 10 Glu residues to 9 Glu residues per 30 residues [24,25]. Similarly, for a constant Glu or Asp residue per 30mer, the replacement of five Val residues by Phe residues causes very large pK_a shifts [26,27], whereas the replacement of the first two Val by two Phe residues per 30 residues has only a very small effect on the pK_a of the Glu or Asp residue (Urry et al., unpublished data) such that going from three Phe residues to five Phe residues, therefore, has a very large effect on the pKa of Glu or Asp. This is referred to as a nonlinear hydrophobic-induced pK_a shift as schematically
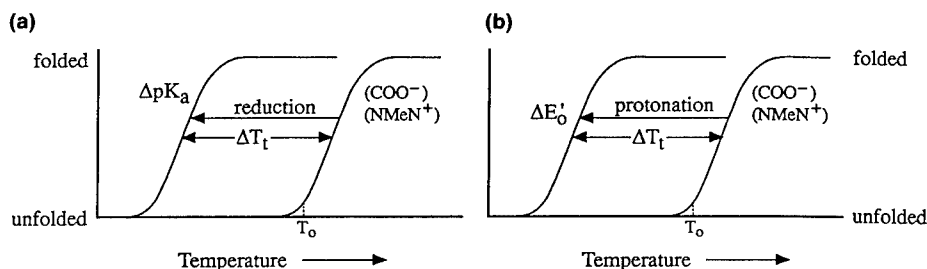
Figure 3.   Design of a model protein containing two different functional groups in their more polar state in which the composition has been adjusted such that the hydrophobic folding transition occurs just above a working temperature, T_0. The energy input that converts either of the functional groups to a more hydrophobic state can drive hydrophobic folding by lowering the value of T_t. The occurrence of the second functional group as part of the hydrophobically folded state changes its property, which could result, e.g., either in the uptake of a proton or an electron.

shown in Figure 4. Now, if the reduction of a single nicotinamide moiety per 30mer occurred with a single Glu or Asp residue per 30mer and there were no Phe residues in the 30mer, then the change in pK_a would be relatively small, as depicted in Figure 4. If, however, there were two or three Phe residues per 30mer and the nicotinamide were reduced, the change in pK_a of the Glu or Asp residue is expected to be much larger, as also shown schematically in Figure 4. This is another aspect of poising and it involves a more efficient conversion of electrical (or light) energy into chemical energy.

Perspectives of Hydrophobic Domain Development and Protein Folding

The preceding formalization of free-energy transduction by means of inverse temperature transitions resulted from a large number of studies on elastic protein-based polymers. Being favorably balanced with hydrophobic valyl side chains and more polar peptide moieties in the backbone such that it exhibited its hydrophobic folding and assembly transition between room temperature and body temperature, the parent model elastic protein, poly(VPGVG), provided a fortunate starting

**Schematic Representation of
Non-linear Hydrophobic-induced pKa Shifts**
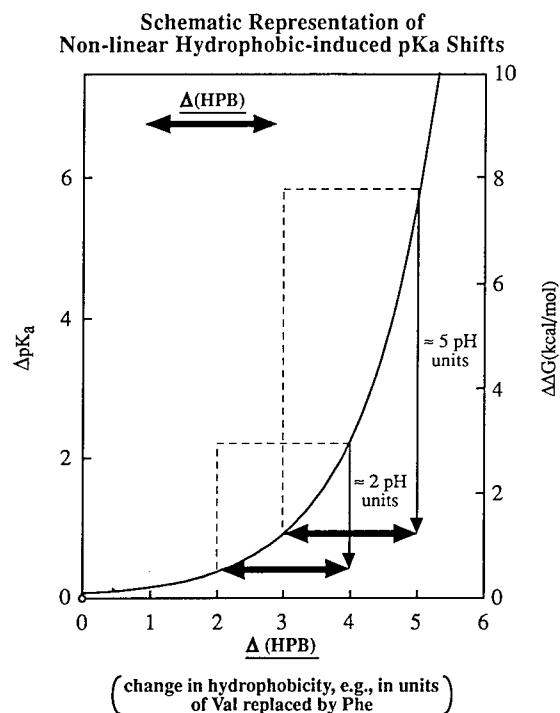
Figure 4. A schematic representation of the dependence of a pK_a shift, a ΔpK_a, on hydrophobicity of the model protein capable of exhibiting a hydrophobic folding and assembly transition. The effect of a particular hydrophobicity change, ⟺, on the pK_a depends in a nonlinear way on the hydrophobicity of the initial state. This is referred to as *poising*.

point for the studies. Being a repeating pentameric molecular structure with one position that would support substitution, it became possible to obtain relevant data on substituents of significance to protein structure and function. The next several paragraphs constitute introduction and correlation of these experimentally derived perspectives with current theoretical and experimental studies on protein folding by consideration of the ingenious, cooperative, "hydrophobic zipper" model of Dill [28], a related, earlier free energy of nucleation arising from hydrophobic contacts facilitated by β-turns due to Matheson and Scheraga [29] and the insightful experimental studies of Dobson et al. [30,31], which combine three experimental approaches to suggest a more generalized hydrophobic collapse as an initial step in protein folding.

The Theoretical Hydrophobic Zipper Model by Dill

The theoretical "hydrophobic zipper" model by Dill [28] is a binary lattice model in which there are two types of units in a polymer chain: hydrophobic units (i_H) and polar units (i_P). What Dill demonstrated with his model is that given sequences can result in cooperative hydrophobic folding. With the correct sequence, the association of a particular pair of hydrophobic units can facilitate the association of another pair to the growing hydrophobic cluster. The zipper can be initiated in a segment of polymer chain, $(i - n) \cdots i \cdots (i + n)$, when a polar moiety i is flanked on both sides by hydrophobic sequences. The location of the polar i moiety identifies the location of a potential turn whereby the next-nearest neighbors ($i - 1$) and ($i + 1$) can associate hydrophobically. Having done so, however, places the next pair of hydrophobic residues, ($i - 2$) and ($i + 2$), in proximity such that they may more readily associate and so on along the zipper resulting in the cooperative formation of a hairpinlike structure held together by a string of hydrophobic contacts.

This theoretical hydrophobic zipper model can now be made more physical by replacing the binary code with a relevant hydrophobicity scale as in Figure 1 and Table II. A real β-turn structure with a 10-atom hydrogen-bonded ring utilizing the C—O of residue i and the NH of residue $i + 3$ can be considered to achieve the folding of the chains as in a hairpin with the resulting formation of a cross-β-structure as in Figure 5. This is closely equivalent to the Matheson and Scheraga analysis [29], with the addition here of the hydrophobicity scale. The particular structure of Figure 5, resulting in an antiparallel alignment of β-chains, is such that hydrophobic association between chains occurs at residues R_i and R_{i+3} (as occurred in part in the development of the hydrophobicity scale of Table II), and then at residues R_{i-2} and R_{i+5}, residues R_{i-4} and R_{i+7}, etc. The resulting structure will be hydrophobic on one side. With a continued folding back and forth, β-barrels could be formed with hydrophobicity inside as occurs in extramembrane proteins or with the hydrophobicity outside as occurs in forming pores or channels in lipid bilayer membranes.

The hydrophobicity scale can be used to form a single-residue hydrophobicity plot for a given primary structure and followed by an odd-and-even sequence editing to identify β-chains with a hydrophobic sidedness as would occur in the hydro-

Hydrophobic Folding of Cross - β - Structures
($C_\alpha - C_\alpha$ virtual bond representation)
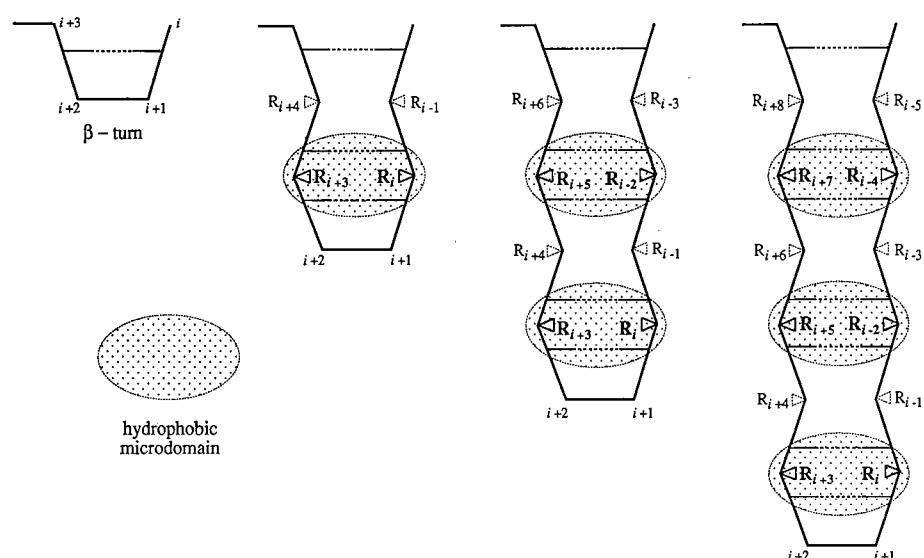
Figure 5. The "hydrophobic zipper" model of Dill put in real structural terms of a β-turn and a cross-β-structure where the hydrophobic side chain pairing occurs with residues i and $i + 3$, $i - 2$, and $i + 5$, etc. The hydrophobicity scale of Table II can be used to compare relative hydrophobicities by summing and normalizing each possible cross-β-structure that could form from a given primary structure. Those structures with the lowest normalized value of T_i would be the most favored.

phobically folded β-barrels of the fibronectin type 3 domains [32,33]. Similar sequence editing could be used to identify amphiphilic α-helices.

The next step toward a more correct physical model is to introduce the interactions between hydrophobic (apolar) and polar groups that are responsible for triggering the energy conversions listed in Table I, i.e., those interactions that control whether hydrophobic folding will occur at a particular temperature. From the studies on the elastic protein-based polymers comes the perspective that a sufficiently polar group can exert such a pull on the surrounding water in achieving its own hydration that it can destructure sufficiently proximal waters of hydrophobic hydration, thereby removing the thermodynamic driving force for hydrophobic folding [1]. Alternatively, in this competition for hydration, it should be appreciated that sufficiently hydrophobic domains can raise the pK_a of a carboxyl group, for example. The important point here is that the cooperative "hydrophobic zipper" model of Dill provides the foundation whereby it becomes possible to think of more physical treatment of cooperative hydrophobic folding in proteins. It might also be appreciated that the cooperative hydrophobic folding to form the β-spiral structure of poly[f_V(VPGVG),f_X(VPGXG)], as has been discussed for many years (see [1]

and references therein) in the terminology of Dill, is an example of a helical hydrophobic zipper.

The Experimental Hen Lysozyme Studies of Dobson

An emerging general perspective from the experimental side of protein folding is seen in the elegant work of Dobson and co-workers on hen lysozyme [30,31]. Its elegance arises from the temporal comparison of stopped-flow far-UV circular dichroism studies, pulsed hydrogen exchange labeling, and fluorescence intensity studies. From the stopped-flow far-UV circular dichroism studies on hen lysozyme, the circular dichroism pattern of the native protein was obtained after 4 ms, indicating that the average relative orientation of peptide backbone moieties was consistent with the native α-helix and β-structure. The pulsed hydrogen-exchange studies indicated essentially no slowing of peptide NH exchange as would be expected of stable α-helix and β-structure, suggesting that the structure was being held together by tertiary structure interactions. The fluorescence intensity studies using native tryptophan residues argued that hydrophobic folding occurred within the first few milliseconds of the refolding process. This caused Dobson and co-workers [30] to suggest that the fundamental question is "the degree to which secondary structure formation in the first steps of folding is accompanied by global structural collapse, for example, by hydrophobic side chain interactions, or indeed whether secondary structure is formed as a consequence of such a collapse" with reference to Dill [28].

Realizing that circular dichroism does not report directly on the hydrogen bonding of secondary structure but rather on the relative orientation of peptide moieties that are associated with the α-helix and β-structures, the studies suggest that hydrophobic folding is an early event that can then stabilize secondary structure. As suggested by Dill [28], this provides an answer to the Levinthal paradox that there is not sufficient time for a protein chain to sample all of the possible torsional angle conformations in the process of finding the global minimum. It suggests that forming energetically favorable hydrophobic domains can be a dominant protein folding process, and as we have reported above, controlling hydrophobic domain formation can be the means with which to achieve all of the energy conversions of which living organisms are capable.

Bibliography

[1] D. W. Urry, Angew. Chem. (German) **105**, 859 (1993); *Ibid.*, Angew. Chem. (English) **32**, 819 (1993).
[2] D. W. Urry, Int. J. Quantum Chem., Quantum Biol. Symp. **14**, 261 (1987).
[3] J. T. Edsall, J. Am. Chem. Soc. **57**, 1506 (1935).
[4] H. S. Frank and M. W. Evans, J. Chem. Phys. **13**, 507 (1945).

[5] W. Kauzmann, Adv. Protein Chem. **14**, 1 (1959).

[6] C. Tanford, in *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* (Wiley, New York, 1980).

[7] A. Ben-Naim, in *Hydrophobic Interactions* (Plenum Press, New York, 1980).

[8] J. T. Edsall and H. A. McKenzie, Adv. Biophys. **16**, 53 (1983).

[9] G. Némethy and H. A. Scheraga, J. Phys. Chem. **66**, 1773 (1962).

[10] G. Némethy and H. A. Scheraga, J. Chem. Phys. **36**, 3382 (1962).

[11] M. V. Stackelberg and H. R. Müller, Naturwissenschaften **38**, 456 (1951).

[12] M. M. Teeter, Proc. Natl. Acad. Sci. U.S.A. **81**, 6014 (1984).

[13] C.-H. Luan, R. D. Harris, K. U. Prasad, and D. W. Urry, Biopolymers **29**, 1699 (1990).

[14] D. W. Urry, C.-H. Luan, R. D. Harris, and K. U. Prasad, Polym. Prep., Am. Chem. Soc. Div. Polym. Chem. **31**, 188 (1990).

[15] C.-H. Luan, T. M. Parker, K. U. Prasad, and D. W. Urry, Biopolymers **31**, 465 (1991).

[16] C.-H. Luan and D. W. Urry, J. Phys. Chem. **95**, 7896 (1991).

[17] D. W. Urry, Prog. Biophys. Mol. Biol. **57**, 23 (1992).

[18] D. W. Urry, M. M. Long, R. D. Harris, and K. U. Prasad, Biopolymers **25**, 1939 1953 (1986).

[19] D. W. Urry, R. D. Harris, M. M. Long, and K. U. Prasad, Int. J. Pept. Protein Res. **28**, 649 (1986).

[20] D. W. Urry, D. C. Gowda, T. M. Parker, C-H Luan, M. C. Reid, C. M. Harris, A. Pattanaik, and R. D. Harris, Biopolymers **32**, 1243 (1992).

[21] D. W. Urry, B. Haynes, Z. Zhang, R. D. Harris, and K. U. Prasad, Proc. Natl. Acad. Sci. U.S.A. **85**, 3407 (1988).

[22] D. W. Urry, L. C. Hayes, T. M. Parker, and R. D. Harris, Chem. Phys. Lett. **201**, 336 (1993).

[23] L. A. Strzegowski, M. B. Martinez, D. C. Gowda, D. W. Urry, and D. A. Tirrell, J. Am. Chem. Soc. **116**, 813 (1994).

[24] D. W. Urry, S. Peng, T. M. Parker, D. C. Gowda, and R. D. Harris, Angew. Chem. (German) **105**, 1523 (1993); *Ibid.,* Angew. Chem., Int. Ed. Engl. **32**, 1440 (1993).

[25] D. W. Urry, S. Peng, and T. M. Parker, J. Am. Chem. Soc. **115**, 7509 (1993).

[26] D. W. Urry, D. C. Gowda, S. Peng, T. M. Parker, and R. D. Harris, J. Am. Chem. Soc. **114**, 8716 (1992).

[27] D. W. Urry, D. C. Gowda, S. Peng, T. M. Parker, N. Jing, and R. D. Harris, Biopolymers **34**, 889 (1994).

[28] K. A. Dill, F. M. Fiebig, and H. S. Chan, Proc. Natl. Acad. Sci. U.S.A. **90**, 1942 1946 (1993).

[29] R. R. Matheson, Jr. and H. A. Scheraga, Macromolecules **11**, 819 (1978).

[30] P. A. Evans and S. E. Radford, Curr. Op. Struct. Biol. **4**, 100 (1994).

[31] C. M. Dobson, P. A. Evans, and S. E. Radford, Trends Biochem. Sci. **19**, 31 (1994).

[32] A. R. Kornblihtt, K. Umezawa, K. Vibe-Pedersen, and F. E. Baralle, EMBO J. **7**, 1755 (1985).

[33] D. J. Leahy, W. A. Hendrickson, I. Aukhil, and H. P. Erickson, Science **258**, 987 (1992).

Ab Initio and Molecular Dynamics Study of Dibenzotricyclic Calcium Antagonists: A Rigid Model Approach

JIAN WANG, AATTO LAAKSONEN*, AND RUSSELL J. BOYD

Department of Chemistry, Dalhousie University, Halifax Nova Scotia, Canada B3H 4J3

RONGSHI LI AND PATRICK S. FARMER

College of Pharmacy, Dalhousie University, Halifax, Nova Scotia, B3H 3J5

Abstract

A new class of calcium antagonists (dibenzotricyclic compounds) is studied by means of reaction field *ab initio* calculations and molecular dynamics simulations. The central ring of these tricyclic molecules is found to be more important to the calcium antagonistic potency than the two phenyl rings. The central ring with antagonistic potency shows hydrophobic character, thus the interaction between the drug and the binding sites is assumed to be dominated by hydrophobic interactions. Variation of the flexure angle, the angle between the two phenyl rings, does not change the hydrophobic property of the central ring significantly, therefore it is not expected to affect the interaction between the drug and binding site directly. The effect of the flexure angle on calcium antagonistic potency, the relation between drug affinity of these tricyclic molecules and their ionization energies, and the interaction of calcium ions with the central ring are discussed. © 1994 John Wiley & Sons, Inc.

Introduction

Calcium ions (Ca^{2+}) play a vital role in many biological processes including a variety of enzymatic reactions, activation of excitable cells, coupling of electrical activation, and cellular secretion, homeostasis. The regulation of the intracellular concentration of this ion makes it possible to control Ca^{2+}-dependent processes. According to the World Health Organization [1], calcium antagonists are a chemically heterogeneous group of compounds that modify Ca^{2+} mobilization or Ca^{2+} action. Verapamil [2], diltiazem [3], and nifedipine [4] are among representatives of the first generation of calcium antagonists, currently used for the treatment of angina and hypertension. Calcium antagonists can affect cardiovascular hemodynamics by three principal actions: coronary arterial dilation, peripheral arterial dilation, and a negative inotropic effect. For example, nifedipine dilates the renal arteries by abolishing autoregulation and it also dilates the cerebral arteries. This widespread vasodilation reduces systemic vascular resistance both in animals and in humans, increasing both contractility and heart rate [5]. It has also been reported

---

* Permanent address: Division of Physical Chemistry, Arrhenius Laboratory, University of Stockholm, S-106 91 Stockholm, Sweden.

that diltiazem may cause serious adverse effects such as cutaneous vasculitis [6], thrombocytopenia [7], heart block [8], parkinsonism [9], and fatal renal and hepatic toxicity [10].

Attempts to find tissue selective calcium antagonists have been made for more than a decade. Recently, a series of compounds with dibenzotricyclic structures (DBTs) was reported as a new class of calcium antagonists having selectivity for cardiac tissue over vascular tissue [11]. These DBTs, for example, I–III in Figure 1, may confer antagonistic activity without a side effect on blood pressure. The structure-activity studies show that calcium antagonistic activity is closely related to the flexure angle, the angle between the planes of the two phenyl rings in DBTs. Most recently, a series of compounds such as dibenzothiazepinone (IV–V in Figure 1) has been synthesized [12]. These compounds are expected to have smaller flexure angles due to a planar amide bond in the central ring of the tricyclic system. Thus enhanced calcium antagonistic activities of these compounds are anticipated.

Due to increasing interest in calcium ion channels and in the interaction between calcium ion-channel related binding sites (receptors) and drugs (ligands), we carried out computational chemical studies of these dibenzotricyclic molecules. The purpose
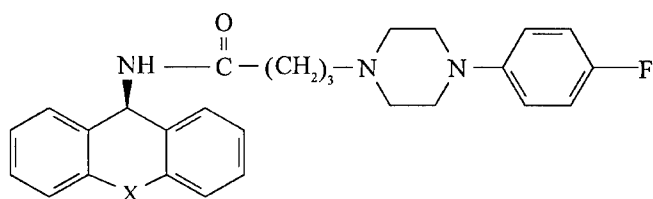
I   (X= CH$_2$O)

II (X= CH$_2$S)

III (X= CH=CH)
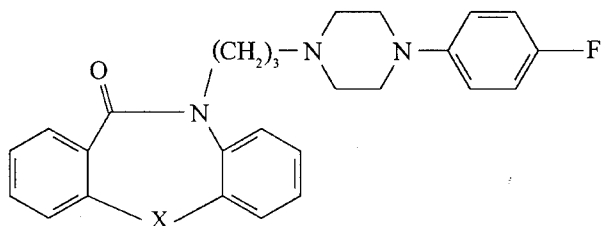
IV (X=O)

V   (X=S)

Figure 1.   The structures of some new dibenzotricyclic calcium antagonists.

of this article is to investigate the conformations and electrostatic energy distributions of some model molecules related to calcium antagonistic molecules. Five model molecules (1–5 in Figure 2) are chosen to model the respective candidate drug molecules shown in Figure 1. *Ab initio* calculation, which includes a reaction field to take into account the solvent effect, was employed in this study. Particular attention was paid to the bridge region of these tricyclic compounds and to their flexure angles. In order to understand the interaction between dibenzotricyclic molecules and water and between the calcium ions and these molecules, molecular dynamics (MD) simulations have been carried out for some model tricyclic molecules in solution. The article is organized as follows. In the second section we briefly describe the computational details of *ab initio* calculations and molecular dynamics simulations. The optimized geometries, atomic charges, electrostatic energies, and flexure angles of these model tricyclic molecules are compared in detail in the third
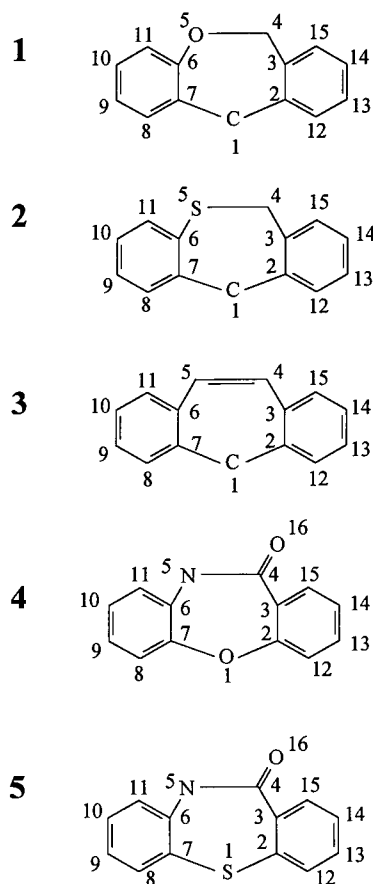
Figure 2. The structures of model molecules used in the present study.

section. The results of the interactions of these model molecules with water and Ca^{2+} ions are discussed in the fourth and fifth sections, respectively.

Computational Details

The conformations, atomic charges, and electrostatic potential were calculated by *ab initio* (Hartree–Fock, HF) methods using the Gaussian 92 program [13]. The geometries were optimized at the HF/STO-3G level using gradient methods. The self-consistent reaction field (SCRF) method was employed in the calculation to allow for the solvent effect on the electronic features of these molecules and to model the "true" molecular environment during geometry optimization. In this approach, the solute is placed in a spherical cavity immersed in a continuous medium with a dielectric constant ε. A molecular dipole will induce a dipole in the medium and in return, the electric field applied to the solute by the solvent reaction dipole will lead to a change of the solute dipole; this interaction is repeated until the system reaches a net stabilization. A dielectric constant of 78.25 is used for the water solvent, and the volume of the spherical cavity of model molecules is defined as the volume inside a contour of 0.001 electron/bohr^3 density. The atomic charges, which are fitted to the electrostatic potentials at points, selected according to the Merz–Singh–Kollman scheme [14], were calculated at the HF/6-31G(d,p) level using the HF/STO-3G optimized geometries. The electrostatic potentials at nuclei are calculated at the same level as well.

Molecular dynamics simulations were carried out for aqueous solutions of molecules 1 through 5. The simulations were performed without and with added $CaCl_2$ salt. All simulations were started by placing two drug molecules in the simulation cell together with 254 water molecules. In simulations with added salt, six randomly chosen water molecules were replaced with two Ca^{2+} and four Cl^- ions. The systems are equilibrated during 50 ps, followed by a production run of 70 ps. The atom–atom radial distribution functions (RDF) have been used to examine the structure of simulated aqueous solutions. The RDF $g(r_{ij})$ describes the spatial distribution of atom j around atom i and gives the probability of finding a pair of atoms ij at the distance r apart, relative to the probability expected for a completely random distribution at the same density, i.e., bulk density, ρ_0 [15]. The integral of $g(r_{ij})$:

$$z(r_{ij}) = 4\pi\rho_0 \int_0^r g(r_{ij}) r^2 \, dr$$

gives the number of atoms j around atom i as a function of the distance r and is called the coordination number (integral) of the RDF.

The drug molecules are kept rigid throughout the simulation. This is, of course, only an approximation but it can be justified by the stiffness of the tricyclic compounds treated in the present study. It is important to maintain the same molecular geometry in the simulations as the one used to calculate the atomic charges and electrostatic potentials. Another reason to maintain a rigid geometry is to study the effects of varying the flexure angle of the drug compounds on intermolecular (between drug and water or between drug and Ca^{2+}) interactions. In a forthcoming

article, we will report MD simulations for compounds **1** through **5** allowing full flexibility and we will investigate the effect of attaching a hydrocarbon chain.

Standard Lennard–Jones atom–atom potentials were used for the drug molecules and the ions. Atomic charges are calculated in the present work and discussed in the next section. Rigid simple point charge (SPC) water [16] molecules were used in all simulations. Standard combination rules were employed to construct the cross-interaction potentials between unlike atoms. The potential parameters are given in Table I. The simulations are carried out in an NVT ensemble at 300 K, using normal densities. The simulation software is a modified version of the McMoldyn package [17].

Ab Initio **Results and Discussions**

Analysis of Atomic Charges and Electrostatic Potentials

The computed atomic charges and electrostatic potentials at heavy atoms for molecules **1**, **2**, **3**, **4a**, **4b**, and **5** are tabulated in Tables II and III, where **4a** corresponds to a HF/STO-3G optimized structure, while **4b** corresponds to an enlarged flexure angle ($124.8°$ and $169.9°$ for **4a** and **4b**, respectively). Replacement of the oxygen atom of **1** by a sulphur atom results in noticeable changes of the charge density distribution in the bridge region. For example, C1 and C2 in **2** have almost doubled the charges they have in **1**, and the charge at C3 decreases from 0.003 in **1** to −0.124 in **2**. Also the C4 and C6 charges are decreased by more than a factor of two from **1** to **2**. Comparing the charge distribution of **3** with **1**, we observe that, like from **1** to **2**, the charges at C2 and C5 are increased while they are decreased at C4 and C6 from **1** to **3**. In contrast to the charge changes in the phenyl rings, the bridge region atoms have in general larger change than the phenyl rings from **1** to **2** or from **1** to **3**. The fact that **2** and **3** represent molecules having greater antagonistic potencies than **1** may indicate a favored interaction of drug and binding sites occurring at the bridge region of these tricyclic molecules, rather than at the phenyl rings as suggested by Kurokawa et al. [11].

Table III lists the electrostatic potential at corresponding nuclei calculated at the HF/6-31G(d,p) level with HF/STO-3G optimized geometries. Relative to **1**, **2** and

TABLE I. Lennard–Jones interaction parameters used in the MD simulation.

	ϵ (kcal/mol)	σ (Å)
C	0.113	3.350
N	0.082	3.310
H	0.113	3.350
O	0.136	2.950
S	0.404	3.520
Ca^{2+}	1.206	4.500
Cl^-	0.118	2.209

TABLE II. Atomic charges at HF/6-31G**//HF/STO-3G level.[a,b]

No. of atom	Model molecules					
	1	**2**	**3**	**4a**	**4b**	**5**
1	C −.369	C −.786(−)	C −.344(+)	O −.325(+)	O −.422(−)	S −.097(+)
2	C .175	C .439(+)	C .197(+)	C .261(+)	C .481(+)	C −.038(−)
3	C .003	C −.124(−)	C .102(+)	C −.099(−)	C −.217(−)	C .073(+)
4	C .089	C .032(−)	C −.220(−)	C .701(+)	C .866(+)	C .666(+)
5	O −.404	S −.280(+)	C −.198(+)	N −.657(−)	N −.838(−)	N −.772(−)
6	C .227	C .007(−)	C .112(−)	C .257(+)	C .418(+)	C .421(+)
7	C .152	C .391(+)	C .181(+)	C .168(+)	C .197(+)	C −.088(−)
8	C −.245	C −.309(−)	C −.254(−)	C −.164(+)	C −.224(+)	C .014(+)
9	C −.171	C −.137(+)	C −.141(+)	C −.202(−)	C −.199(−)	C −.285(−)
10	C −.157	C −.153(+)	C −.161(−)	C −.123(+)	C −.123(+)	C −.078(+)
11	C −.208	C −.188(+)	C −.212(−)	C −.261(−)	C −.339(−)	C −.319(−)
12	C −.224	C −.322(−)	C −.255(−)	C −.219(+)	C −.368(−)	C −.002(+)
13	C −.161	C −.117(+)	C −.144(+)	C −.104(+)	C −.026(+)	C −.208(−)
14	C −.136	C −.157(−)	C −.155(−)	C −.156(−)	C −.205(−)	C −.074(+)
15	C −.195	C −.193(+)	C −.205(−)	C −.143(+)	C −.150(+)	C −.229(−)
16				O −.569	O −.619	O −.556

[a] Atomic charges were calculated by fitting to the electrostatic potentials at points selected according to Merz–Singh–Kollman scheme [14].

[b] The "+" sign in the parentheses indicates greater atomic charges with respect to that of atoms at respective positions of **1**, while the "−" sign corresponds to a smaller value relative to that of **1**.

3 have the same qualitative change of the electrostatic potential at the C2, C4, C6, C9, and C10 positions. The same conclusion applies to the atomic charges except the trend for C10. This implies that the change of charges and electrostatic potentials at these positions can be used as one of the indices to discuss the antagonistic activity of some new candidate molecules that have similar structures, such as **1–3**.

The newly synthesized compounds dibenzoxazepinone and dibenzothiazepinone [12] are modelled by **4a** and **5**, respectively. The atomic charges and electrostatic potentials at the heavy atoms of these two model molecules are listed in Tables II and III. Relative to **1**, **4a** and **5** have in general opposite changes of charges and electrostatic potentials at the positions of C2, C4, C6, and C9 as **2** and **3** do. For instance, **4a** and **5** correspond to increased charges and electrostatic potentials at C4 and C6 positions with respect to that of **1**, whereas **2** and **3** have relatively decreased values at the same positions in comparison with **1**. There are increased charges and decreased electrostatic potentials at C9 of **5** and **6** relative to **1**, respectively, but the converse holds at the C9 position of **2** and **3**. These opposite changes in charges and electrostatic potentials indicate that **4a** and **5** will not have as great Ca^{2+} antagonistic potencies as **2** and **3**, if **4a** and **5** are assumed to interact with the same binding sites as **2** and **3**. In our previous study [18], we found that **3** corresponds to a positive electrostatic potential near the bridge region, whereas **4a** and **5** correspond to a negative electrostatic potential in the same region, and

TABLE III. HF/6-31G**//HF/STO-3G electrostatic potential at nuclei (a.u.).[a]

No. of atom	Model molecules					
	1	2	3	4a	4b	5
1	C −14.72099	C −14.71344(+)	C −14.72313(−)	O −22.26231(−)	O −22.26263(−)	S −59.19464(−)
2	C −14.71778	C −14.71784(−)	C −14.72164(−)	C −14.63468(+)	C −14.61869(+)	C −14.67184(+)
3	C −14.71763	C −14.71750(+)	C −14.72236(−)	C −14.69683(+)	C −14.70392(+(	C −14.69343(+)
4	C −14.66069	C −14.68966(−)	C −14.72719(−)	C −14.57829(+)	C −14.58283(+)	C −14.58002(+)
5	O −22.29391	S −59.21319(−)	C −14.72723(+)	N −18.29658(+)	N −18.29095(+)	N −18.30012(+)
6	C −14.67271	C −14.69197(−)	C −14.72243(−)	C −14.66156(+)	C −14.65312(+)	C −14.66049(+)
7	C −14.72170	C −14.71063(+)	C −14.72172(−)	C −14.64796(+)	C −14.62590(+)	C −14.68427(+)
8	C −14.72697	C −14.72554(+)	C −14.73201(−)	C −14.71045(+)	C −14.70675(+)	C −14.70746(+)
9	C −14.72964	C −14.73003(−)	C −14.73000(−)	C −14.71443(+)	C −14.71412(+)	C −14.71534(+)
10	C −14.72751	C −14.72956(−)	C −14.73242(−)	C −14.71133(+)	C −14.71579(+)	C −14.70776(+)
11	C −14.72994	C −14.72978(+)	C −14.73097(−)	C −14.70940(+)	C −14.71322(+)	C −14.71102(+)
12	C −14.73017	C −14.72617(+)	C −14.73198(−)	C −14.71364(+)	C −14.71524(+)	C −14.71084(+)
13	C −14.72889	C −14.72560(+)	C −14.72994(−)	C −14.70314(+)	C −14.70588(+)	C −14.70662(+)
14	C −14.72797	C −14.72748(+)	C −14.73239(−)	C −14.71688(+)	C −14.72626(+)	C −14.71412(+)
15	C −14.72800	C −14.72699(+)	C −14.73092(−)	C −14.70766(+)	C −14.71747(+)	C −14.71076(+)
16				O −22.33308	O −22.34231(+)	O −22.33240

[a] The "+" sign in the parentheses indicates greater electrostatic potential with respect to that of atoms at respective positions of 1, while the "−" sign corresponds to a smaller value relative to that of 1.

therefore, if molecules **4a** and **5** do exhibit calcium antagonistic potency experi-
mentally, they probably interact with different binding sites or with different states
of the binding sites of **2** or **3**.

Flexure Angle Effects

In a recent structure-activity study [11] it was found that the flexure angles between
the two phenyl rings in these tricyclic molecules are important to the Ca^{2+} antag-
onistic activity. The HF/STO-3G SCRF optimizations yield flexure angles of 120.6,
122.5, 114.5, 124.8, and 112.6° for molecules **1**, **2**, **3**, **4a**, and **5**, respectively.
Model molecule **2** has a larger angle than **1** but **3** has a smaller angle than **1**. Since
2 and **3** represent molecules having more potent antagonistic activities than **1**, the
flexure angle does not directly affect the chemical interaction between the drug and
binding site of the receptor. The effect of the flexure angle may be due to a specific
structural requirement of receptors to ligands or to the interaction of drugs with
other molecules while approaching the binding sites. We will examine the flexure
angle effect on the interaction between drug and water in detail in the third section.

The Affinities of Drug Molecules

The interactions between calcium antagonists and their binding sites are chemical
interactions [19]. One of the important indices to measure the activity of a drug
is its affinity. The affinity of a drug molecule is closely related to its ionization
energy or its electron affinity. Table IV summarizes the values calculated at the
HF/6-31G(d,p) and HF/STO-3G levels with HF/STO-3G optimized geometries for
the model molecules **1**, **2**, **3**, **4a**, and **5**, respectively. The ionization energies and
electron affinities were estimated by means of Koopmans' theorem [20], which,
of course, is an approximation, but nonetheless useful for qualitative discussions.

The HF/6-31G(d,p) and HF/STO-3G calculations yield similar orders of ioniza-
tion energies and electron affinities. Model molecules **2** and **3**, which have greater
antagonistic potencies, correspond to smaller ionization energies than **1**, while the
electron affinities of both **2** and **3** are less than that of **1**. In view of the positive

TABLE IV.   Approximate ionization energies and electron affinities at
the HF/STO-3G and HF/6-31G(d,p) levels (a.u.).[a]

	HF/STO-3G		HF/6-31G(d,p)	
	I	A	I	A
1	0.26059	0.25764	0.31337	0.13360
2	0.23363	0.21752	0.28927	0.10623
3	0.21681	0.25676	0.29018	0.13403
4a	0.24763	0.21214	0.32143	0.09032
5	0.24217	0.21614	0.32496	0.08688

[a] All calculations used HF/STO-3G optimized geometries.

electrostatic potential near the bridge region [18], we suggest that these tricyclic molecules interact with the binding sites by donating electrons into the sites. Thus, the ionization energy, which measures the energy required to remove an electron from a molecule, can be used to measure the affinity of a drug molecule. Therefore, molecules **2** and **3** have a greater drug affinity than **1** due to their smaller ionization energies. Even though the ionization energies of **4a** and **5** are smaller than that of **1**, they are still greater than the corresponding values of **2** and **3**. Consequently, we predict that **4a** and **5** will have less antagonistic potencies than **2** and **3**. Of course, this prediction is under the assumption that they interact with the same binding sites as **2** and **3**, as we already pointed out in the previous sections.

The Interaction of Drug Molecules with Water

Since water is the natural environment in a biological system, hydration and other specific interactions between drugs and water are of importance in determining the action of drugs in the human body. Furthermore this information can provide insight into the interaction between the drug molecules and their binding sites. Molecular dynamics simulation results are reported, focusing on the interaction between water and the bridge region of these model molecules. The structure of water around the central ring was examined using radial distribution functions: $g(X - O_w)$, where X refers to the atoms of the central ring, and O_w to water oxygen. The coordination numbers (integrals) of the RDF were also included in the figures with the same line pattern corresponding to the RDF.

Figure 3 displays radial distribution functions of the water oxygen around the C1, C4, and O5 atoms of the central ring of **1**. The shape of $g(C1-O_w)$ clearly
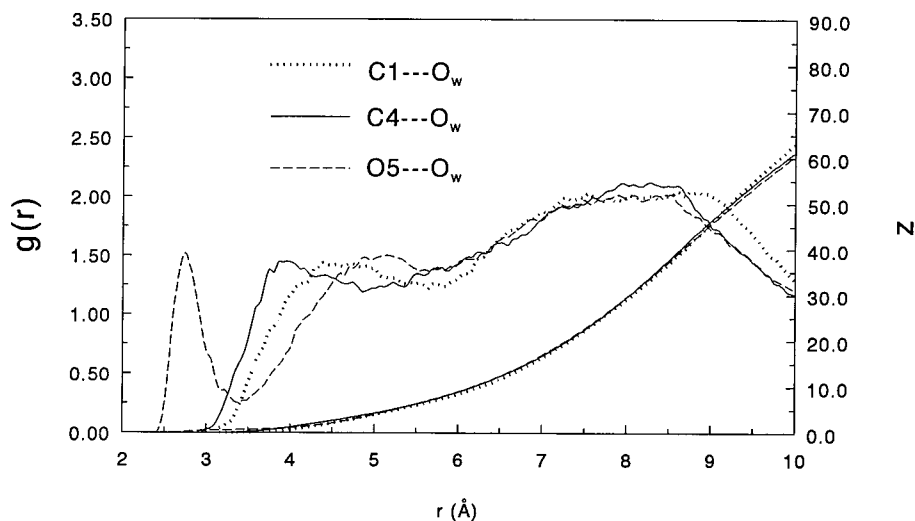
Figure 3.   The radial distribution functions and corresponding coordination numbers of water oxygen with bridge atoms of **1**.
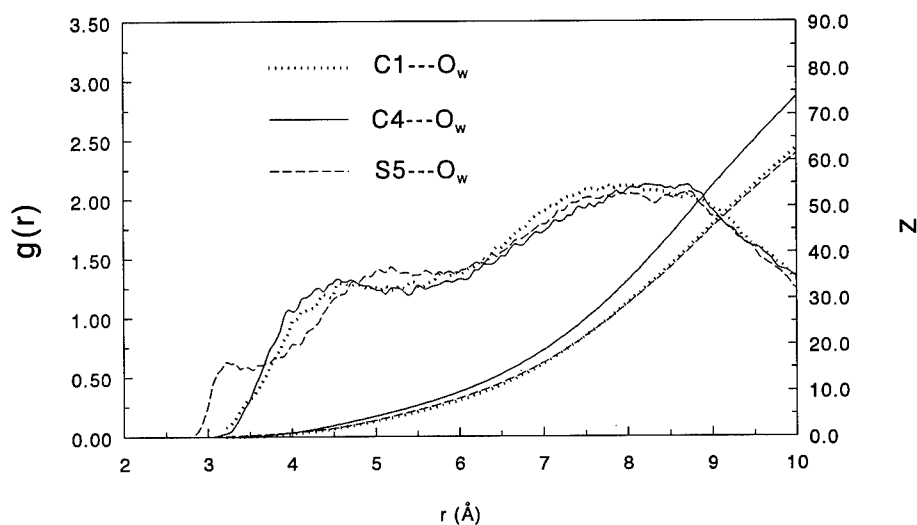
Figure 4.   The radial distribution functions and corresponding coordination numbers of
water oxygen with bridge atoms of **2**.

indicates that the interaction between water and C1 is hydrophobic. The same
observation holds for the interaction between water and C4. On the other hand,
the peak at 2.8 Å in $g(O5-O_w)$ exhibits a weak hydrogen bond between O5 and
water (the dashed line), its corresponding coordination number is 0.66. Replacement
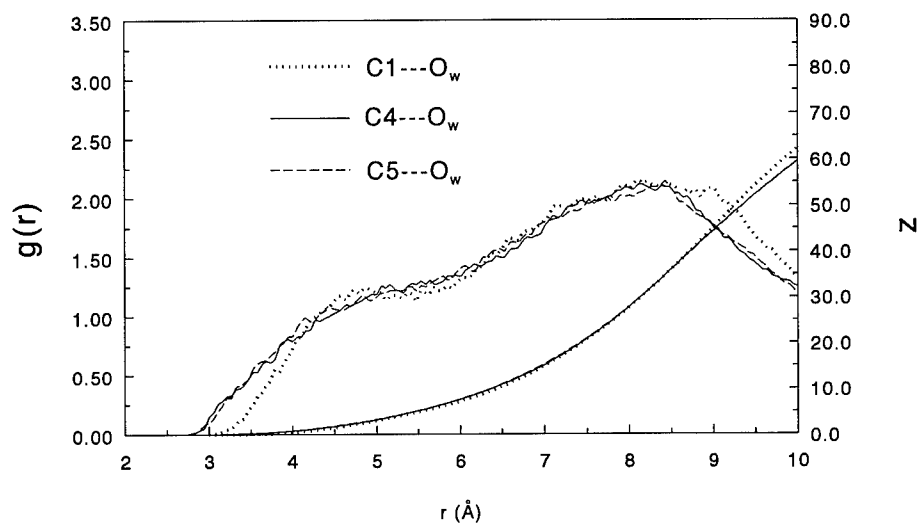
Figure 5.   The radial distribution functions and corresponding coordination numbers of
water oxygen with bridge atoms of **3**.

of the oxygen in the central ring of **1** by sulphur gives rise to a hydrophobic central ring, as illustrated in Figure 4. The same conclusion holds for **3**. Figure 5 illustrates the corresponding radial distribution functions for **3**. Overall, the curve shapes of Figure 5 are closer to Figure 4 than to Figure 3. This indicates that the hydrophobic bridge region in these tricyclic compounds will enhance the antagonistic activities of the drug. It also suggests that the interaction between these tricyclic molecules and their binding sites is dominated by hydrophobic interaction, as suggested by Triggle et al. in the case of dihydropyridine [19].

We have also carried out MD simulations on model molecules **4a**, **4b**, and **5**. Figure 6 is the related radial distribution functions for **4a** and **4b**, respectively. The small peak at 2.8 Å of $g(O1-O_w)$ (dotted line in Figure 6) indicates that there is weak hydrogen bonding between O1 of **4a** and water due to the electronegative oxygen in the bridge region. Although the nitrogen in the bridge region does not appear to have a strong interaction with water (cf. solid line in Figure 6), the O16 of **4a** has a strong interaction with water through hydrogen bonding. With reference to the previous discussion in this section about binding sites favoring a hydrophobic bridge region in tricyclic molecules, **4a** is unlikely to show stronger antagonistic potency than **2** or **3** to the same binding sites. The same conclusion applies to model molecule **5** as well.

It is interesting to see how the flexure angle between the two phenyl rings affects the interaction between water and model drug molecules. Sterically, a smaller flexure angle will more or less block water from entering the bridge region, thus a weaker interaction is expected. Figure 7 shows the radial distribution functions of water oxygen with the O1, N4, and O16 atoms of **4b**. The only difference between **4a** and **4b** is the flexure angle: 124.8° vs. 169.9°. Surprisingly, we did not observe an
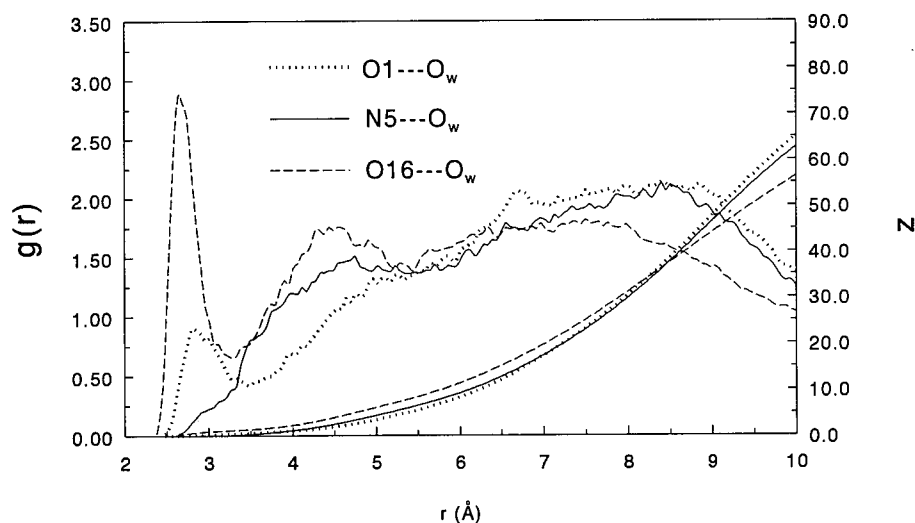
Figure 6.   The radial distribution functions and corresponding coordination numbers of water oxygen with bridge atoms of **4a**.
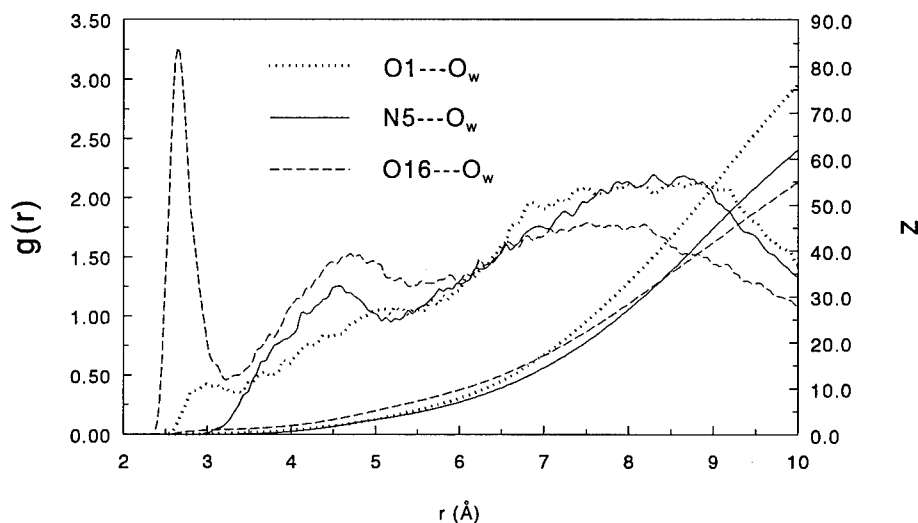
Figure 7.   The radial distribution functions and corresponding coordination numbers of
water oxygen with bridge atoms of **4b**.

enhanced interaction between water and O1 or between water and N4 atoms in
4b. Instead, the hydrogen bond between water and O1 of **4a** is slightly weaker in
4b due to the enlarged flexure angle (Figure 7 vs. Figure 6). Also, nitrogen and
water interactions become weaker as indicated by the reduced intensity of the radial
distribution function and the closest distance in the RDF (cf. Fig. 7). The interaction
between water and O16 is slightly reduced as well after the flexure angle is enlarged.
The coordination numbers corresponding to $g(\mathrm{O}16\text{-}\mathrm{O}_w)$ (dashed lines in Figures
6 and 7) are 1.14 and 1.06 for **4a** and **4b**, respectively. By comparing Figures 6
and 7, we note that in general an enlarged flexure angle does not change significantly
the hydrophilicity of the bridge region in the model drug molecules, and therefore
it does not affect the interaction between the drug molecules and their binding sites
significantly. This suggests that the effect of the flexure angle on Ca^{2+} antagonistic
activity is mainly due to steric blocking of drug molecules entering the bind-
ing sites.

The Interaction between Calcium Ions and Drug Molecules

Of particular significance to the calcium antagonist study is whether or not the
antagonists interact with calcium ions. So far there is no clear experimental evidence,
although there have been some reports of the Ca^{2+} dependence of the drugs' binding
to the calcium channels [21,22]. The allosteric interaction between different classes
of calcium channel antagonists at the dihydropyridine receptor also seems to be
Ca^{2+}-dependent [21]. Recently, Ananthanarayanan et al. [23] studied the inter-
action of calcium channel antagonists (diltiazem) with calcium by means of NMR
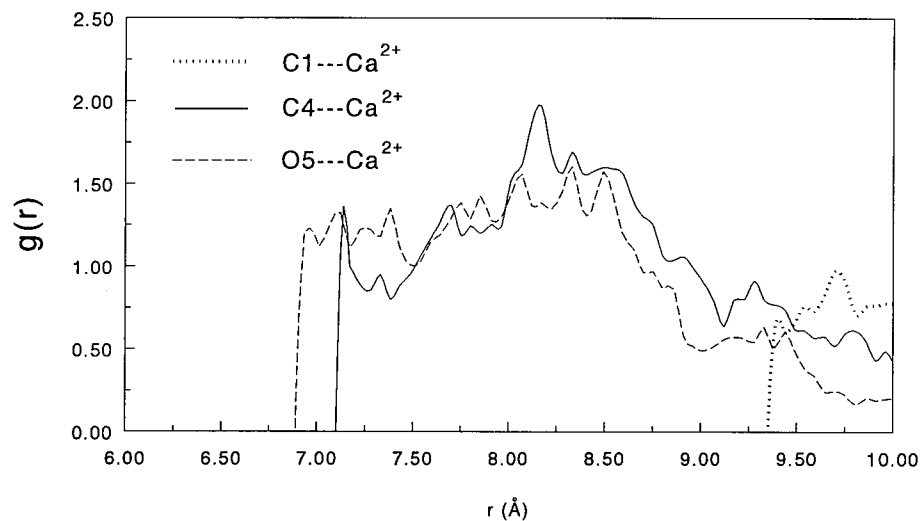
Figure 8. The radial distribution functions of calcium ion with bridge atoms of **1**.

spectroscopy and molecular modeling. We present here our preliminary results based on molecular dynamics simulations.

Similar to the previous section, RDFs are used to examine the probability of finding Ca^{2+} around the model drug molecule. Figures 8 and 9 show the radial distribution functions of Ca^{2+} ions with the bridge region atoms for model molecules
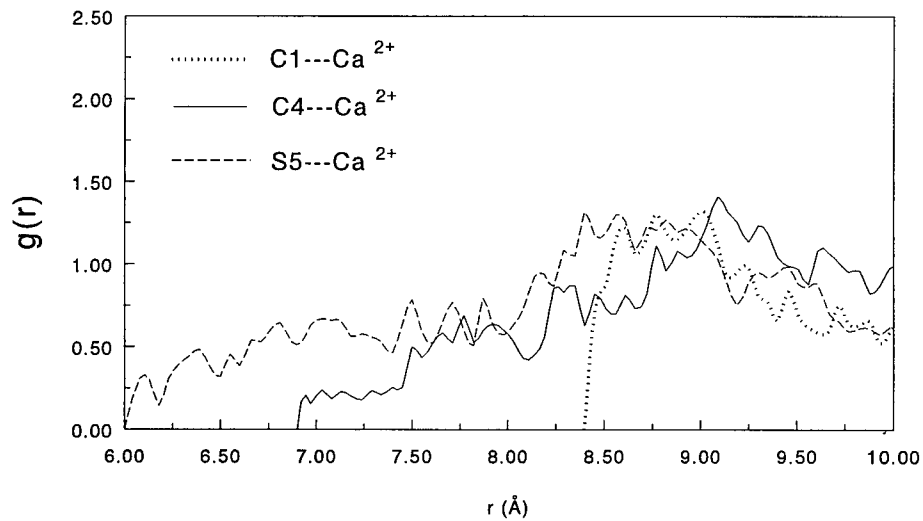
Figure 9. The radial distribution functions of calcium ion with bridge atoms of **2**.

1 and **2**, respectively. We should point out that although the curves in these figures are not smooth enough due to the relatively short running time (140 ps), some information still can be extracted from them.

The shapes of $g(C1–Ca^{2+})$, $g(C4–Ca^{2+})$, and $g(O5–Ca^{2+})$ indicate that there is no direct interaction between Ca^{2+} and model molecule **1**. The "contact" distances in the RDFs are approximately 9.4, 7.1, and 6.9 Å, respectively. Peaks near 8.3 Å can be expected in the long run for the $C1–Ca^{2+}$ and $C4–Ca^{2+}$ pairs. A peak that appears near 9.8 corresponds to the $O5–Ca^{2+}$ pair. This is attributed to Ca^{2+} interacting with the drug central ring through water, i.e., a strongly hydrated Ca^{2+} surrounded by water interacts with the hydrated drug. In addition, the sharp increase of RDFs from the starting distances in Figure 8 manifests the Ca^{2+} unable to come closer to the bridge ring of the drug molecule because of the positive electrostatic potential near the region. So the bridge region is "hard" to Ca^{2+} ions. However, replacement of oxygen in the central ring of **1** by sulfur causes molecule **2** to become "soft" to the Ca^{2+} ions. The contact distances of RDFs in the model molecule **2** (Fig. 9) are shorter than the corresponding distances in model molecule **1**. Similar observations were found in the case of **3**. On the basis of these preliminary results, we propose that there is no direct interaction between Ca^{2+} ions and the central ring of these model drug molecules.

Conclusions

In this work a new class of calcium antagonists (dibenzotricyclic compounds) has been studied by means of reaction field *ab initio* calculations and molecular dynamics simulations within the rigid model approach. Relatively greater changes in atomic charges and electrostatic potentials are found in the central ring than the two phenyl rings of the model tricyclic drug molecules, hence the central ring of these molecules may be more important to the calcium antagonistic potency than the two phenyl rings. The molecular dynamics study of the interaction between the central ring of the drug molecule and water indicates that a central ring with great antagonistic potency has mainly hydrophobic character, and therefore, the interaction between the drug and binding sites may be dominated by hydrophobic interactions. Variation of the flexure angles does not change the hydrophobic property of the central ring significantly, and therefore it is not expected to affect the interaction between the drug and binding site directly. The effect of the flexure angle on calcium antagonistic potency may be due to the specific structural requirement of binding sites to the ligand or to the interaction with other molecules when the drug molecule approaches the receptor. The drug affinity of these tricyclic molecules can be closely related to their ionization energies; the drug action may require donation of an electron from the drug molecule to the binding site. The calcium ions do not coordinate or interact directly with the central ring of the model drug molecules, but rather they interact through the water.

Acknowledgments

Bibliography

[1] P. M. Vanhoutte, Am. J. Cardiol. **59**, 3A (1987).
[2] S. Baky, in *New Drug Annual: Cardiovascular Drugs,* A. Scriabine, Ed. (Raven, New York, 1984), Vol. 2, pp. 71–102.
[3] M. Chaffman and R. N. Brogden, Drugs **29**, 387 (1985).
[4] E. M. Sorkin, S. P. Clissold, and R. N. Brogden, Drugs **30**, 182 (1985).
[5] A. Schwartz, *Calcium Antagonists: New Data on Relationship Between Pharmacology and Clinical Effect,* monograph, 1992; pp. 1–12.
[6] R. A. Sheehan-Date and M. J. D. Goodfield, Postgrad. Med. J. **64**, 467 (1988).
[7] M. Lahav and R. Arav, Ann. Intern. Med. **110**, 327 (1989).
[8] P. C. Waller and W. H. W. Inman, Lancet **1**, 617 (1989).
[9] R. S. Dick and S. S. Barold, Am. J. Med. **87**, 95 (1989).
[10] H. Shalleross, S. P. G. Padley, M. J. Glynn, and D. D. Gibbs, Br. Med. J. **295**, 1236 (1987).
[11] M. Kurokawa, F. Sato, I. Fujiwara, N. Hatano, Y. Honda, T. Yoshida, S. Naruto, J. Mastumoto, and H. Uno, J. Med. Chem. **34**, 927 (1991).
[12] R. Li, P. S. Farmer, M. A. Quilliam, and S. E. Howlett, Drug Des. Discovery **10**, 331 (1993).
[13] M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. DeFrees, J. Baker, J. J. P. Stewart, and J. A. Pople, *Gaussian 92* (Gaussian, Inc., Pittsburgh, PA, 1992).
[14] (a) U. C. Singh and P. A. Kollman, J. Comput. Chem. **5**, 129 (1984); (b) B. H. Besler, K. M. Merz, Jr., and P. A. Kollman, J. Comput. Chem. **11**, 431 (1990).
[15] M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Oxford Science Publications, Oxford, England, 1987).
[16] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces,* B. Pullman, Ed. (Reidel, Dordrecht, Holland, 1981).
[17] A. Laaksonen, Comput. Phys. Commun. **42**, 271 (1986).
[18] R. Li, P. S. Farmer, J. Wang, R. J. Boyd, T. S. Carmeron, M. A. Quilliam, J. A. Walters, and S.E. Howlett, Drug Des. Discovery (in press).
[19] D. J. Triggle, D. A. Langs, and R. A. Janis, Med. Res. Rev. **9**, 123 (1989).
[20] T. Koopmans, Physica **1**, 104 (1934).
[21] R. A. Janis and D. J. Triggle, in *Calcium Channel: Their Properties, Functions, Regulations, and Clinical Relevance,* L. Hurtwitz, L. D. Partridge, and J. K. Leach Eds. (CRC Press, Boca Raton, FL, 1992), p. 197.
[22] H. Schoemaker and S. Z. Langer, J. Pharmacol. Exp. Ther. **248**, 710 (1989).
[23] V. S. Ananthanarayanan, S. Tetreault, and A. Saint-Jean, J. Med. Chem. **36**, 1324 (1993).

Received March 11, 1994

Proton Release Pathway in Bacteriorhodopsin: Molecular Dynamics and Electrostatic Calculations

C. SCHARNAGL,* J. HETTENKOFER, AND S. F. FISCHER

Institut für Theoretische Physik der Technischen Universität München, James-Franck-Strasse, D-85747 Garching, Germany

Abstract

We use molecular dynamics, electrostatic, and quantumchemical calculations to discuss chromophore and protein structural changes as well as proton transfer pathways in the first half of the bacteriorhodopsin photocycle. A model for the molecular mechanism is presented, which accounts for the complex pH dependence of the proton release and uptake pattern found for the M intermediates. The results suggest that transient transfer of the Schiff base proton to a nearby tightly bound water molecule is the primary step, which is accompanyied by dissipation of free energy to the protein. From there, the energetically most favorable proton transfer is to aspartate D85. Arginine R82 is involved in the protein reorientation switch, which catalyzes the pK_a reduction of glutamate E204. This residue is, therefore, identified as extracellular proton release group whose acid base equilibrium regulates the pH-dependent splitting of the photocycle. © 1994 John Wiley & Sons, Inc.

Introduction

Bacteriorhodopsin (BR) is a light-driven proton-pumping protein in the cell membrane of *Halobacterium halobium*. The photoinduced isomerization of the Schiff base linked retinal chromophore (RSB) from all-*trans* to 13-*cis* (for recent reviews see, e.g., [1–3]) initiates a sequence of thermal reactions with an overall turnover time of about 10 ms.

The photocycle consists of spectroscopically distinct principal intermediate states, labeled J, K, L, M, N, and O. Several models account for the complex time courses of their rise and decay, which introduce either parallel photocycles with slightly different, unidirectional kinetic steps (e.g., [4–7]) or one photocycle with reversible reactions between the intermediates (e.g., [8,9]). By the time the L state is reached the change in the retinal geometry is communicated to the protein. The following proton translocations are based on directed sequential alterations in the pK_a's of the retinal Schiff base and strategically located titratable residues. Site directed mutagenesis (e.g., [10–17]) have revealed several key amino acids which are involved in proton transfer and in maintaining the functional state of the pigment. These studies, together with a moderate resolution structure for BR [18] have led to the conclusion, that the isomerization is followed by transfer of the RSB proton to aspartate D85, which is part of a complex counterion near the extracellular protein

---

* To whom correspondence should be addressed.

surface, consisting of two aspartates (D85, D212), an arginine (R82), two tyrosines (Y57, Y185), and bound water molecules [14,19,20].

During the lifetime of the M intermediate with the deprotonated retinal Schiff base two kinetically different steps occur: One is a protein conformational change which guarantees that the reaction proceeds mainly in the forward direction (the "switch step" [8,21–23]); the other step involves pH-dependent proton release to the extracellular and proton uptake from the cytoplasmic side of the membrane [24–26]. The groups which accomplish these proton exchange reactions at the surface must have suitable pK_a's. The proton release complex has initially a high pK_a, which is decreased during the first half of the photocycle. The pK_a of this group in the M intermediate(s) (determined to be approx. 6 [24]) causes the diverging of the photocycle into two pH-dependent alternative pathways: If the pH is lower than the pK_a of this release group, proton release is delayed after uptake. At higher pH a transient proton deficit in the protein develops.

On the cytoplasmic side aspartate D96 and several water molecules are part of the proton uptake channel, which is involved in the reprotonation of the Schiff base in the M → N step [11,27]. The original pK_a of the uptake groups is near 11 [26]. The corresponding difference in the pK_a's of the release and the uptake group is enough to pump protons against an electrochemical gradient of ≈ 300 mV. Proton uptake from the bulk, reprotonation of D96, reisomerization of the retinal Schiff base to all-*trans* and protein relaxations reestablish the original situation [2,25,26].

In this study we discuss the proton release pathway in the bacteriorhodopsin photocycle. The various aspects of the proton transfer mechanism are investigated using a combination of different theoretical methods:

- The retinal isomerization reaction, accompanying protein structural changes and charge separations, is discussed by means of quantum chemical and molecular dynamic calculations. As we start from the medium resolution structure for BR [18], this includes the completion of the structural information due to hydration of the intrahelical regions.
- General electrostatic concepts allow to analyze the energetics of the proton pump in terms of the changed molecular structure during the photocycle. This includes the determination of the pK_a's for different sites with a continuum dielectric model and atomic details. The calculations include pK_a shifts due to the electrostatic effects of burying charged groups in the low dielectric membrane, interactions to protein and water dipoles, and the coupling of titratable residues with each other.

We present a sequence of isomerization driven protein conformational changes which leads to a description of the proton release pathway consistent with experimental results. In particular, we discuss the role of a bound water molecule in the counterion complex as transient proton binding site and its involvement in the dissipation of free energy in the L to M transition as well as the molecular nature of the reorientation switch during the lifetime of the M intermediates. We propose a direct involvement of arginine R82 in the protein reorientation step (M_1 → M_2). Its movement towards the extracellular side catalyzes the pK_a reduction of a

glutamate residue (E204), which acts as proton release group. Our calculations explain the complex pH dependence of the proton uptake and release pattern found for the M intermediates and suggest the pH-dependent splitting of the photocycle at the stage of the M_2 intermediate.

Methods

Energetics of Proton Transfer Reactions

The use of general electrostatic concepts [28–30] allows to analyze and describe proton transfer reactions in terms of the underlying molecular structure. The energetics and the rates can be estimated from the actual polarity of the relevant sites. A certain proton configuration (m) of the protein with N titratable residues will be described by the protonation vector $\vec{x}(m) = (x_1^{(m)}, \ldots, x_N^{(m)})$. Proton transfer between two residues i to j connects two protein states, e.g.,

$$(m) = (\ldots, x_i^{(m)} = 1, \ldots, x_j^{(m)} = 0, \ldots) \rightarrow (n)$$
$$= (\ldots, x_i^{(n)} = 0, \ldots, x_j^{(n)} = 1, \ldots) .$$

The main factor, which determines the activation barrier $\Delta G^{\#}$ (and, therefore, the rate constant), is the corresponding free energy difference between reactant and product state $\Delta G(m \rightarrow n)$. This free energy difference can be correlated with the pK_a difference for proton donor and acceptor:

$$\Delta G^{\#} \propto \Delta G(m \rightarrow n) \simeq -2.3kT[pK_a^{(n)}(j) - pK_a^{(m)}(i)] = \Delta G_{PT} .$$

The free energy for a protonation state of the protein at a given pH [30–32] is

$$\Delta G(m) = \sum_{i=1}^{N} \{x_i^{(m)}(\Delta G_{int}(i) + 2.3kT\,pH)\} + \tfrac{1}{2} \sum_{i,j=1}^{N} W_{i,j}(q_i^0 + x_i^{(m)})(q_j^0 + x_j^{(m)})$$

with

q^0 = charge of the unprotonated residue (acid: -1; base: 0),

$\Delta G_{int}(i)$ = self-energy of residue i in the protein, where all other sites are in their neutral state or enter as fixed background charges,

$W_{i,j}$ = interaction between titrating sites, work required to charge site j in the presence of a charge on site i,

k = Boltzmann's constant,

T = temperature.

The self energy $\Delta G_{int}(i)$ can be further divided up into three contributions:

$$\Delta G_{int}(i) = \Delta G_{model}(i) + \Delta\Delta G_{solv}(i) + \Delta\Delta G_{back}(i)$$

with

$\Delta G_{model}(i)$ = protonation free energy for the residue in aqueous solvent (as reference state), which can be calculated form the pK_a value in solvent: $\Delta G_{model}(i)$ $= -2.3kT\,pK_{a,\,model}(i)$,

$\Delta\Delta G_{solv}(i)$ = difference in reaction field energy for protonated and unprotonated residue in the reference state and the protein,

$\Delta\Delta G_{back}(i)$ = difference in electrostatic interaction energy for protonated and un-protonated residue with permanent dipoles and constant background charges in reference state and protein.

Changes in self energy and site–site couplings shift the acid–base equilibrium for a residue in a given protonation state of the protein, which is conveniently expressed as the actual pK_a value of the residue:

$$pK_a^{(m)}(i) = pK_{int}(i) - \sum_{j=1}^{N} W_{i,j}(q_j^0 + x_j^{(m)})/2.3kT \qquad (j \neq i)$$

with $pK_{int}(i) = -\Delta G_{int}(i)/2.3kT$.

The average protonation of a site i is given by the Boltzmann weighted sum over all 2^N possible protonation states of the protein ($m = 1, \ldots, 2^N$):

$$\langle x_i \rangle = \sum_m x_i^{(m)} \exp[-\Delta G(m)/kT] \Big/ \sum_m \exp[-\Delta G(m)/kT] \,.$$

In this approximation the changes in protonation free energies are assumed to result only from a change in the enthalpy.

Electrostatic Calculations

Calculations of reaction field energies and electrostatic interactions between charged groups were carried out with the finite difference Poisson–Boltzmann (FDPB) method using the program package DelPhi [31,33–35]. Partial charges and van der Waals radii for the amino acids were taken from the CHARMm21 parameter set [36]. The partial charges for the free base and the protonated Schiff base retinal chromophore were calculated with the program package MOPAC 6.0 using the AM1 Hamiltonian [37].

The protein and the membrane region were treated as low dielectric cavity (dielectric constant ε_{in} = 4) embedded in an aqueous medium of dielectric constant ε_{out} = 80 with an electrolyte of ionic strength 0.15 M and an ion exclusion radius of 2 Å. The cytoplasmic surface of the membrane is presumed to be even with glutamic acid E166, the extracellular surface is even with E74. This results in a membrane of approx. 45 Å thickness [38]. Twofold focusing resulted in a final resolution of 1.4 grid/Å. Test calculations using rotational averaging revealed maximal errors for $\Delta\Delta G_{back}$ of 10%, for $W_{i,j}$ between 5% and 10%.

As it is assumed that the aqueous medium contains mobile ions, our calculations account to a certain degree for the difference between bulk pH and surface pH. Due to the omission of the interhelical loop regions and the membrane lipid head groups, the description of the screening will be incomplete.

The pK_a values for titratable amino acids (asp, glu, lys, arg, tyr) in aqueous medium were taken from [39]. Water and R—OH residues (thr, ser) have two states of protonation and correspondingly two pK_a values. For the hydroxyl form

$(R\!-\!OH \rightleftharpoons RO^- + H^+)$ we used $pK_a(ROH) = 15.7$ [40], for the hydronium form $(R\!-\!OH_2^+ \rightleftharpoons ROH + H^+)$ we used $pK_a(ROH_2^+) = -1.7$ [40]. The possible protonation of a backbone peptide group was investigated with the model compound acetamide ($pK_{model} = 0.0$ [40]). The pK_{model} value used for the all-*trans* retinal Schiff base (7.0) was taken from Ref. [32].

As was proposed earlier [41], the *trans–cis* isomerization of the C13$=$C14 bond induces additional twists in neighboring single bonds, which can result in a stereochemical pK_a reduction. To determine these pK_a shifts for the 13-*cis* intermediates, we calculated quantum chemically (MOPAC 6.0 [37], AM1 Hamiltonian, lowest 100 singly and doubly excited configurations) the total energies for the isolated neutral and protonated chromophore, whose overall geometry was kept near the value found in the energy minimized protein structure (see below) by constraining the dihedrals. The energy differences relative to the all-*trans* chromophores enter as modifies pK_{model} values for the RSB.

Preparation of the Bacteriorhodopsin Structure

The medium resolution structure for bacteriorhodopsin, determined by electron cryomicroscopy [18], has been taken as starting point for our investigation of bacteriorhodopsin and its intermediates. Energy minimization and molecular dynamics calculations were carried out with the CHARMm force field ([36], version 21 parameter set, dielectric constant $\varepsilon = 1$). Coordinates for polar hydrogen atoms were built using the HBUILD facility of CHARMm. Partial charges for the Schiff base have been determined using the program package MOPAC 6.0 [37]. The torsional and bond stretch force constants and equilibrium values for the polyenlike, unprotonated retinal were taken from the CHARMm parameters. For the protonated retinal Schiff base, these values were scaled to reflect the quantum chemical determined [41] bond features.

The protonation states for titratable groups which participate in proton transfer were taken to be in accordance with experimental observations (e.g., [10–17]) or a recent theoretical calculation of pK_a values [32]—protonated: RSB, D115, D96, Y57, Y185, T89, R82, E204; deprotonated: D85, D212). The ionization state for other titratable residues was selected according to their protonation at pH 7. The electrostatic calculation [32] also supports the repositioning of R82 from the extracellular side to the interior of the protein.

The structure of Henderson et al. [18] describes only the helical part of bacteriorhodopsin; no coordinates for the atoms in the interhelical loop regions are provided. We refrained from the explicit construction of the missing loop regions and adopted heavy patch residues for the terminal groups of the helices. The integrity of the protein was maintained by applying harmonic constraints, which increased continuously from an inner 10 Å radius cyclindrical region without constraints to the outer region [force constant 0.2 kcal/(mol Å^2)]. This selection of harmonic restoring forces simulates the protein–membrane interaction and ensures that all atoms fluctuate approximate with the same mean amplitude. Several molecular dynamics calculations [42,43] revealed that the Henderson structure [18] is close

to a minimum energy conformation. The application of constraints to keep the outer surface of the protein intact imposes, therefore, no unphysical restrictions.

The positions of intramembrane water molecules (TIP3 water model [44]) were carefully inspected: To use the experimental information about water cavities in the native state, no vacuum geometry optimization was performed. In a first step, each residue with a solvent accessible surface $> 10 \text{ Å}^2$ was surrounded by an equilibrated 8 Å water sphere. Water molecules with bad contacts (<2.8 Å distance to protein heavy atoms) were eliminated. The positions of the solvent molecules were optimized relative to fixed protein atoms. The following short (3 ps) molecular dynamics simulation (rapid heating from 0 K to 300 K) was used to create a random set of conformations as input for the geometry optimization algorithm utilizing the threshold accepting principle [45]. This algorithm allows large displacements of the atoms from the starting configurations and supports the inspection of a large region of the potential energy surface and leads, therefore, to a structure near to the global minimum. Fourfold application of this solvation–optimization procedure led to a water saturated structure. Subsequent optimization of solvent and protein and a 160 ps molecular dynamic calculation (10 ps heating to 300 K, 50 ps equilibration, 100 ps free dynamics at 300 K, graduated harmonic constraints applied) helps to identify tightly bound water molecules.

Generation of Intermediates

Experimental information about the light-induced changes in the chromophore geometry (e.g., [2,3,21]) argues for an all-*trans* to 13-*cis* isomerization as primary photochemical event. The potential which governs the motion of a dihedral angle φ in CHARMm is $V_\varphi = K_\varphi[1 + \cos(n\varphi + \delta)]$ with $n = 2$ (periodicity) and $\delta = 180$ (phase factor) in the case of the planar, all-*trans* retinal chain conformation. Quantum chemical [3,46] calculations and fs time resolved spectroscopy [2,3] suggest that the excited state potential minimum is near the ground state isomerization barrier. We use the information from these theoretical calculations to construct an approximate S1-potential surface for the C13=C14 dihedral in the retinal Schiff base: $V_\varphi^* = K_\varphi^*[1 + \cos(n\varphi)]$ with $K_\varphi^* = 5.6$ kcal/mol and the periodicity $n = 2$. The phase factor $\delta = 0$ shifts the S1-equilibrium positions to $\pm90°$. The excited state topology was completed with a new set of partial charges for the protonated retinal resulting from a quantum chemical calculation (INDO-SDTCI [47]). The photoisomerization was induced by abruptly changing the C13=C14 torsional potential from S0- to S1-topology and placing therefore the bond to the excited state maximum. A short molecular dynamics calculation monitored the relaxation of the chromophore away from the Franck–Condon region to the new equilibrium position, which was completed in 270 fs—a value also found experimentally [2]. The isomerization direction is determined by the interaction of the chromophore with the surrounding protein. Strong electrostatic interaction between the Schiff base proton and the negative partial charge of the main chain oxygen at position 212 in helix G favors the twist from $180°$ to $-90°$.

To enforce the completion of the all-*trans* → 13-*cis* isomerization on the ground state potential surface, we assumed the storage of 50 kcal/mol of torsional energy

in the C13=C14 bond and varied the equilibrium position of the dihedral angle in 30° steps to the final 13-*cis* value (0°). As the isomerization time is short [2] compared to protein relaxation times, the equalized geometry of the K intermediate was optained with a frozen protein. Further relaxation of the 13-*cis* chromophore transfers energy from the reaction coordinate to other degrees of freedom. Subsequent protein structural changes lead to the L intermediate. Different molecular models were discussed for the proton transfer and the M intermediates.

To test the plausibility of the generated structures, we performed quantum chemical calculations of the absorption spectra using the INDO-SDTCI approach [47]. The inclusion of singly, doubly, and triply excited electron configurations in the configuration interaction procedure accounts for the polarizabilities in the ground and excited states which are essential for the electronic structure of the chromophore. The retinal, charged, and polar groups (including water) in a 10 Å surrounding spherical region were treated as supermolecule. This approach allows one to consider the charge induced wavelength shifts. The calculated values (λ_{max}/nm; dipole strength/Debye^2) are BR(547; 145) − J(625; 54) − L(524; 148).

Results and Discussion

The BR Ground State

Structure: As the Henderson structure [18] contains only detailed information about the location of the β-ionone ring, the position of the conjugated portion of the chromophore has to be worked out. It is determined by the steric strain exerted from the tightly packed residues in the vicinity (tryptophanes W86, W182, W189, proline P186) and the electrostatic interactions between the protonated Schiff base and anionic (aspartates D85, D212), cationic (arginine R82) or polar (tyrosines Y57, Y185 and water) groups. The equilibrated all-*trans* chromophore is almost planar except a slight twist of the C14—C15 bond (dihedral −165°). Small twist around single bonds were also found experimentally [48].

It is well established [14,20] that the primary proton acceptor in the first half of the photocycle (D85) is part of a complex counterion system that consists of the aspartate residues D85, D212, arginine R82, the tyrosines Y57 and Y185, and probably a tightly bound water molecule [14,19]. The quadrupolelike arrangement of RSB, R82, D85, and D212, which was proposed by NMR spectroscopy [20], is a constituent part of our BR structure (see Fig. 1). The direct counterion for the protonated RSB is D212 which is bound to the same helix as the RSB and is, therefore, constrained to stay in its vicinity. Compared to the original structure, we find a displacement of D85, which moves 2 Å away from the Schiff base, due to electrostatic repulsion between the two aspartic acids. The gap is filled by a tightly bound water molecule (named X1 in the structure), which can now serve as direct counterion in the proton release pathway.

It is experimentally established [12,49,50], that water molecules are functionally important, especially in the uptake of the proton from the cytoplasm. Water structural changes during the photocycle have been observed [51]. Neutron diffraction investigations [52] suggest that there are at least four tightly bound water molecules
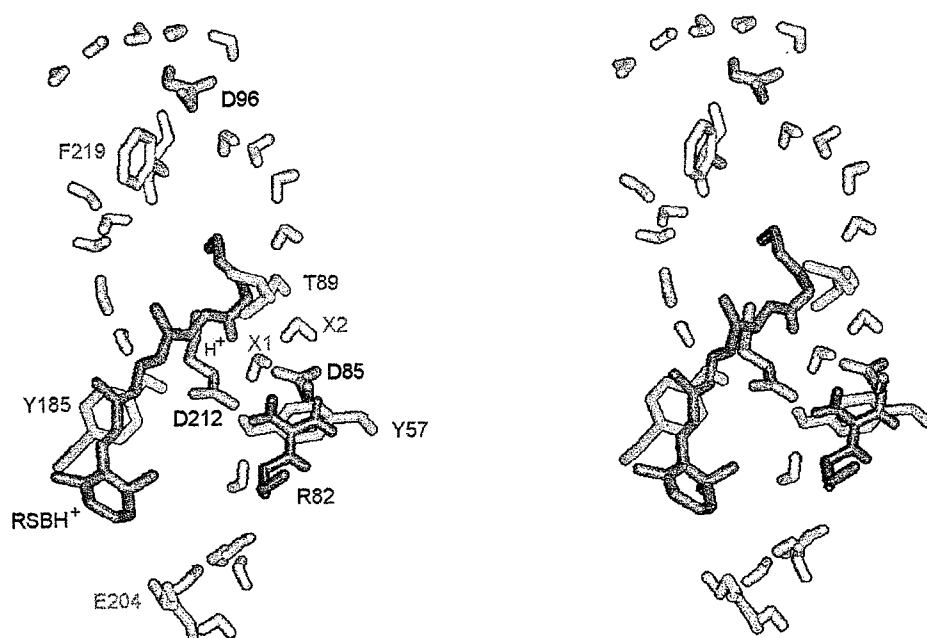
Figure 1.   BR intermediate—stereoview of retinal and functionally important residues in the transmembrane region (amino acids are labeled by their one-letter code, water molecules are labeled by X).

near the Schiff base and several exchangeable protons participating in the pathway across the membrane. Our BR structure shows two water chains in the proposed proton conduction channel on the cytoplasmic side, which may serve as proton wires for the transfer to the Schiff base. One spans the region from aspartate D96 over threonine T89—two residues with an essential role in proton uptake [11,12,16,54]—to another tightly bound water (named X2 in Fig. 1) and aspartate D85, the primary proton acceptor in the release pathway. The other chain is interrupted by a hydrophobic region near phenylalanine F219 and ends near the tyrosines Y57 and Y185. Recent experimental results [50] indicate that at least 15 mol of water are required for optimal proton transfer to and from the Schiff base. This is a much larger amount than needed for a single chain of water molecules between RSB and D96. It probably represents diffuse hydration of the interhelical region on the cytoplasmic side of the protein in spite of the hydophobic nature of the proton channel itself.

   Acid–Base Equilibria: The usefulness and limitations of pK_a calculations for the BR ground state based on the medium resolution structure [18] was already shown [32]. Our calculations differ by this investigation through the inclusion of water molecules in the structure, which appear either as permanent dipoles in the background term or are treated explicitly as titrating groups, which can accept or donate protons [35].

We include only a subset of titratable groups in our calculations: Besides the RSB, we regard members of the counterion complex (D85, D212, R82, Y57, Y185, water X1), four groups in the uptake channel (D96, T89, water X2, and D115), and E204 in the extracellular region. As our C13=C14 isomerization model revealed strong electrostatic interaction between the Schiff base proton and a main chain oxygen, we examined the importance of this peptide group as transient proton acceptor. All other ionizable groups were included in a fixed protonation state (corresponding to bulk pH 7), assuming, that there is no large pK_a shift compared to the solvent value. As a consequence, our titration calculations are reliable only in the pH range from 3 (protonation of carboxylates) to 11 (deprotonation of tyrosines and lysines).

Table I shows the contributions of solvation and background interaction term to pK_a shifts in the BR ground state. As found for other membrane systems [32,35,53], the pK_{int} value is determined by the desolvation penalty for burying charged groups in the low dielectric inner membrane region and favors the neutral protonation state. The shift is more pronounced for groups with a localized charge distribution (water, tyrosine, threonine) while it is less important for residues with largely delocalized charge distributions (especially arginine and the Schiff base). The interaction with nontitratable residues and protein dipoles shows the tendency to reduces the unfavorable shifts due to desolvation for members of the counterion complex. Although D96 and E204 are surrounded by water molecules, the background term calculated for these residues does not compensate the unfavorable solvation term. The solvent molecules around these residues are arranged according to the neutral form of the aspartate or glutamate groups. As for the electrostatic calculation the orientation of background dipoles had to be kept fix, our calculations do not account for the stabilization of the anionic charge form due to a changed orientation of the water dipoles. For example, the magnitude of the unfavorable

TABLE I.   Contributions of solvation and background terms to the pK_a shift in the energy minimized structure of bacteriorhodopsin (BR).

	pK_{model}	ΔpK_{solv}	ΔpK_{back}	$\langle x_i \rangle$	pK_a
RSB	7.0	−4.3	2.7[3.5]	1	15.2
D212	3.9	11.0	−8.5	0	−3.8
D85	3.9	11.2	−2.9[−0.9]	0	−2.3
R82	12.5	−10.1	0.9[−4.9]	1	26.1[21.7]
X1	−1.7	−19.0	−0.4	0	−10.5
Y57	10.1	14.6	−4.0	1	34.2
Y185	10.1	14.6	0.9[1.6]	1	30.2
X2	−1.7	−19.0	1.1[1.7]	0	−15.4
T89	15.7	14.1	0.1[1.1]	1	32.3
E204	4.3	11.4	4.7	1	20.7
D96	3.9	11.7	3.2	1	18.6

Included are the average degree of protonation $\langle x_i \rangle$ of each residue and the final pK_a-value in the ground state configuration. Values in square brackets are for the OD1-HD tautomer for D85.

contribution of the water molecules to the charged form of E204 is approximately +8 pK_a units.

Table II shows the site–site coupling for members of the counterion complex. Due to its extended, alternating charge distribution the interactions between the protonated Schiff base and members of the counterion complex is balanced out. The largest couplings are local pairwise interactions between R82–D85, D212–Y57, and X2–T89. For the carboxylate and guanidyl groups of D85, D212, and R82, we encounter the problem of hydrogen-tautomerism: Which oxygen should receive the proton, which nitrogen should be the donor? In contrast to other attempts [32], which neutralize or distribute the charges symmetrically, we discuss both possibilities explicitly for the important residue D85. The two tautomers correspond to the two physical situations, where the interaction to arginine R82 (OD1-HD tautomer, included in square brackets in Tables I and II) or the interaction to the Schiff base RSB and the water molecules X2 and X1 dominates (OD2-HD tautomer). The same differentiation holds for D212: The OD1-HD tautomer (adopted for the calculation) optimizes the interaction to Y57 and the RSB, the other tautomer to Y185.

To gain some insight into the effect of structural fluctuations on the results, we calculated the pairwise electrostatic interactions for the 2800 frames of our 100 ps molecular dynamics run. This is facilitated by the fact that the screening of electrostatic interactions due to induced surface charges is not prominent in this membrane bound protein. Therefore, we determined the effective dielectric constant ε_{eff} for the interaction term $W_{i,j}$ in the static structure as described earlier [34]:

$$\varepsilon_{eff} = W_{i,j}(\varepsilon = 1, \text{Coulomb})/W_{i,j}(\varepsilon_{in}, \varepsilon_{out}, \text{FDPB}) .$$

$W_{i,j}(\varepsilon_{in}, \varepsilon_{out}, \text{FDPB})$ is the interaction term calculated with the FDPB method, where the membrane is treated as low dielectric region surrounded by the high dielectric solvent, $W_{i,j}(\varepsilon = 1, \text{Coulomb})$ is the corresponding value calculated with Coulomb's law and a constant dielectric constant. The ε_{eff} values vary between $0.9*\varepsilon_{in}$ and $1.5*\varepsilon_{in}$ and indicate no large dielectric inhomogeneities in the protein. From the time series for the $W_{i,j}(\varepsilon = 1, \text{Coulomb})$ values, mean values, and standard deviations can be extracted (lower lines in Table II). The largest fluctuations in the order of 1–2 pH units are found for the OH groups of water, tyrosine, and threonine residues. No value for the static structure lies outside the distribution. It becomes also clear that the strong interaction of D85 and R82 dominates over the interaction between D85 and X1. Fluctuations in the background term ΔpK_{back}, calculated with the same approach, are approximately 1 pK unit for residues in the active site and 2 pK units for the groups of E204 and D96.

The calculated average degrees of protonation $\langle x_i \rangle$ for the residues (Table I) confirm the experimental determined protonation states and are constant over the pH range from 3 to 11. The pK_a values for the residues in this ground-state configuration are also included in Table I. Their maximal uncertainties are ± 3 pK units. The available corresponding experimental values are: 13.2 (RSB [55]), <2.5 (D85, D212 [10,56,57]), >12 (Y57, Y185 [32]), 13.6 (R82 [15]), and >11 (D96 [26]). The general overestimation of pK_a shifts by the applied method has its origin

TABLE II. Site–site coupling in bacteriorhodopsin.

$W_{ij}/2.3$ kT	D212	D85	R82	X1	Y57	Y185	X2	T89
RSB	8.9	6.8[6.1]	5.3	11.4	5.3	8.6	7.3	8.0
	(10.2 ± 0.3)	(6.4 ± 0.3)	(6.6 ± 0.3)	(8.7 ± 1.0)	(5.8 ± 0.4)	(8.2 ± 0.3)	(6.7 ± 0.7)	(7.1 ± 0.4)
D212		9.1[8.7]	11.0	20.4	23.2	12.8	7.1	6.0
		(8.8 ± 0.3)	(14.3 ± 0.6)	(15.2 ± 1.3)	(20.6 ± 1.8)	(13.1 ± 0.5)	(7.5 ± 1.4)	(5.9 ± 0.3)
D85			16.9[18.4]	15.2[13.3]	7.5[8.0]	6.0[5.3]	16.0[16.5]	11.0[10.0]
			(20.2 ± 0.8)	(16.0 ± 1.1)	(7.4 ± 0.4)	(5.7 ± 0.2)	(15.0 ± 1.1)	(11.3 ± 1.8)
R82				13.3	11.9	5.6	9.5	6.5
				(15.3 ± 2.0)	(13.5 ± 0.8)	(7.7 ± 0.4)	(10.0 ± 1.8)	(6.7 ± 0.4)
X1					7.4	7.9	14.2	10.0
					(6.4 ± 1.1)	(6.4 ± 0.6)	(15.5 ± 1.9)	(11.5 ± 2.4)
Y57						4.0	3.7	2.8
						(4.1 ± 0.2)	(4.1 ± 0.8)	(2.8 ± 0.1)
Y185							4.9	5.0
							(4.7 ± 0.4)	(4.7 ± 0.2)
X2								22.7
								(19.5 ± 3.8)

The first line of each item reports values for the energy minimized BR structure (values in square brackets are for the OD1-HD-tautomer of D85). The second line includes mean values and standard deviations calculated from a 100 ps dynamics simulation.

in the forced constant geometry for the charged and neutral group, which prevents the residue from changing position in the protein to optimize the electrostatic interaction [32,31].

As in our approximation pK_a shifts are only due to mutual Coulombic interactions of the groups, the site–site coupling values W_{ij} (Table II) give the contribution for each residue i due to the pairwise interaction to the other sites j. For the RSB we find that the positive charge on R82 lowers the pK_a by 5 units, the negative charges on D85 and D212 raise the value by 7 and 9 units, respectively. This means that both aspartate residues are necessary to hold the retinal in its protonated form over a large pH range. The corresponding experimental values from mutation experiments [15] are a reduction by 2.5 units due to R82 and an increase by 5 units due to D85. The dominant contribution in lowering the pK_a for D85 is the positive charge on R82, as the rise due to the negative D212 compensates widely the reduction due to the positive RSB. The corresponding mutation experiment [15] reveals a reduction by 4.5 units due to R82.

Under the assumption of a fixed geometry, it is also possible to analyze how the proton affinities change if some of the groups are neutralized. Switching from the ground state proton configuration $(m) = (\ldots, x_i = 0(1), \ldots)$ to a configuration $(m') = (\ldots, x_i = 1(0), \ldots)$ shifts the pK_a's for the other residues proportional to the change in electrostatic potential due to group i. These values can be compared to the results of mutation experiments (e.g., [15,17]). For the situation where R82 is neutralized (e.g., R82Q mutant), the removal of a positive charge leads to an increase of the pK_a values for the RSB (+5 units), D85 (+16 units), and D212 (+11 units). The considerably lower experimental value for aspartate D85 (+4.5 units [15,17]) is a clear hint for severe structural reorientations occurring upon this mutation. If both D85 and R82 are uncharged, the calculated pK_a reduction for the RSB by 2 units fits to the experimentally found reduction from 13.2 to 10.6 [17]. If R82 is charged and D85 is neutral, the spectroscopic titration experiment finds a reduction of pK_a (RSB) from 13.2 to 8.4 and the appearance of a second component with an apparent pK_a of 10.6 [15]. Our results indicate that this charge distribution in the protein is coupled to a drastic reduction in pK_a for the RSB (-7.2 units), D212 (-9 units), T89 (-10 units), D115 (-5 units), and R82 (-10 units). The second component could, therefore, be due to the deprotonation of R82 or T89 in the pH range around 10.

Ground-State Charge Heterogeneities: Deviations from the BR ground-state proton distribution appear at extreme low pH values, where either D85 or D212 protonate. This is experimentally established by the purple–blue transition of the membrane and the blocking of the proton pump activity, which was attributed to protonation of D85 at acidic pH [14,56,57].

Small shifts and amplitude changes in the spectrum of the unphotolyzed wild-type bacteriorhodopsin with increasing pH above 8 suggest some pH-dependent heterogeneity [58]. There are at least three additional states: a blue shifted (λ_{max} 480 nm) all-*trans* component, a N-like species possibly with deprotonated D96 [59] and a red-shifted species which is attributed to the ionization of a tyrosine residue [60]. Two different BR states (α, β in a pH-dependent population ratio)

which are different with respect to the conformation (ionization) of at least one specific side chain were postulated from kinetics experiments [4–7]. It was shown that these components undergo different photocycles: In the α-state of BR, proton transfer from the Schiff base is catalyzed, while the β-state does not support or even impedes this reaction step.

The discussion of heterogeneous charge distributions in bacteriorhodopsin is possible, if we regard the proton configurations, which are energetically near the ground-state configuration. One obvious origin of a charge heterogeneity near the chromophore is the formation of a salt bridge between R82 and D85, which needs ca. 7 kcal/mol under the assumption of optimal orientation (OD1-HD tautomer of D85). If this salt bridge is maintained during the isomerization, the preferential proton acceptor site will be blocked for a subset of bacteriorhodopsin molecules. In contrast to the protonation of D85 at low pH, which shifts the absorption wavelength 35 nm bathochromically, the salt bridge would induce only a small wavelength shift as judged from the 10 nm red shift found for the double mutant R82A/D85N [15]. The role of R82 in the ground state heterogeneity was experimentally established [17] by the fact that for the R82A mutant the redshift of the chromophore absorption found for wild type bacteriorhodopsin around pH 9 did not occur. This investigation [17] revealed also the involvement of a transient protonation of D85 on the rate of thermal *trans–cis* isomerization of the retinal chromophore. The changed charge distribution near the chromophore will, therefore, induce some structural heterogeneity, too. The other type of heterogeneity, initiated by the pH catalyzed deprotonation of a residue, can be discussed in terms of the response of the other residues to the ejection of a proton: the deprotonation of D115 in BR, e.g., changes in proton affinities for the residues in the active site according to the individual pairwise interactions W_{ij}(ΔpK$_a \approx$ +3.5 units).

If the acid–base equilibration of the amino acid occurs faster than the formation of all of the photocycle intermediates, only one cycle should be observed [7] and the inhomogeneity will be averaged out in the kinetics. However, more than one parallel cycle is expected if some of the residues that affect the rate of formation of an intermediate have protonation/deprotonation rates on a comparable or slower time scale. From the definition of the acid–base equilibrium constant $K_a = 10^{-pK_a} = k_{dissoc}/k_{assoc}$ the deprotonation time can be estimated [7]. Assuming that the proton association with a pK$_a$ value between 2 and 9 is diffusion controlled with $k_{assoc} = 10^{11}$ M^{-1} s^{-1} in aqueous medium (10^6 M^{-1} s^{-1} in nonaqueous systems), one calculates deprotonation times that range from 10^{-9} s to 10^3 s.

The L Intermediate

For the discussion of molecular mechanisms for proton transfer reactions in bacteriorhodopsin, the L structure is the key intermediate. The proton transport is based on directed changes in the pK$_a$'s of the Schiff base and specific groups in the protein. In the L species all, or nearly all, of the acquired free energy (experimental values vary between 10 kcal/mol (calorimetric measurements [61]) and 23 kcal/mol (photoacustic data [62])) will be manifested in the changed pK$_a$ values of the

groups, which participate in the proton transfer. The changes in pK_a can be accomplished by several conditions: (A) The distortion of the retinal geometry could decrease the pK_a of the RSB [41]. (B) A reduction in its pK_a can be the result of a change in the strength and geometry of hydrogen bonds. (C) Changed electrostatic interactions, e.g., repulsion to positively charged residues and inhomogeneous electric fields in the binding pocket [7] induce a high sensitivity of the electrostatic interactions to even small protein conformations changes during the photocycle.

Structure: The structure of the L intermediate is shown in Figure 2. The relaxed 13-*cis*-chromophore exhibits large-to-moderate twists in the dihedrals neighboring the isomerized $C13 = C14$ bond ($C12 - C13$: 153°, $C14 - C15$: 170°, $C15 = NZ$: 157°), which indicates that the torsional strain caused by the double bond isomerization is not localized to a particular neighboring bond as it is assumed by di-*cis* models (e.g., [63]). The introduction of twists in the retinal moiety close to the Schiff base was also characterized by FTIR spectroscopy [48]. These twists give rise to a calculated stereochemical pK_a reduction for the retinal chromophore by 2 pK units, which enters as reduced $pK_{model} = 5$ for the 13-*cis*-chromophore, compared to $pK_{model} = 7$ for the all-*trans* retinal.

The accompanying reorientation of the Schiff base proton towards helix G disrupts the approximate quadrupolelike arrangement of RSB, D85, D212, and R82 which controls the electrostatic interactions in the BR ground state. The dominant interactions of the protonated RSB in the L structure are strong hydrogen bonds to the D212 peptide and carboxylate group (nearest distance to NZ 3.0 Å and 3.4 Å, resp.). The large attraction of the Schiff base proton to the D212 main chain oxygen even determines the isomerization direction along the excited state potential surface (see
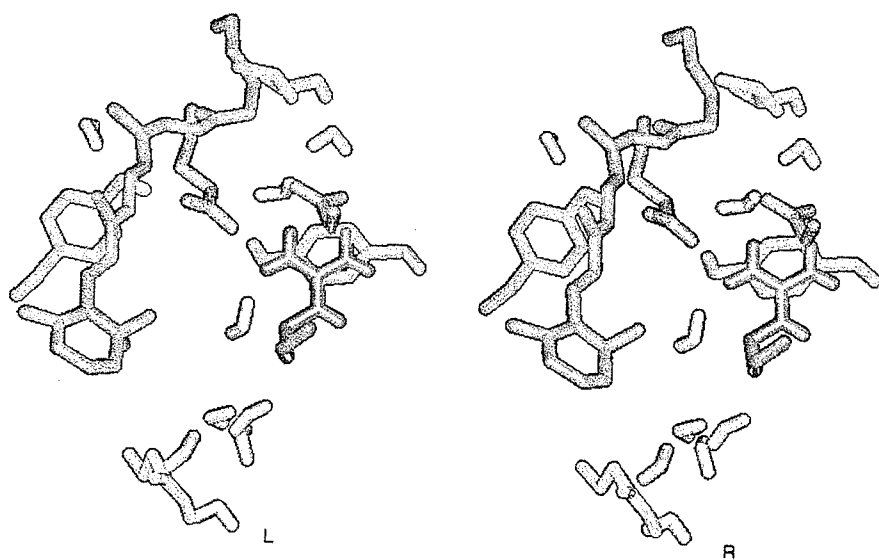
Figure 2.   L intermediate—stereoview of retinal and functionally important residues in the proton release pathway.

methods). An increased hydrogen-bond strength of the Schiff base was also found experimentally by Raman spectroscopy [2,21]. The isomerization induced strain in the binding pocket manifests itself experimentally by backbone conformational changes which give rise to the changed pattern in the amide I/II-bands in FTIR difference spectra (e.g., [9,22]), which appears even if the RSB will not be deprotonated in the following step. Another prominent change occurring upon L formation is the observed stronger hydrogen bonding of water and O — H side chains of the protein (e.g., threonine, tyrosine) [27,51,54].

The deformation of the binding pocket upon isomerization induces the movement of threonine T89 (on helix C) and tyrosine Y57 (on helix B) against each other. As a consequence, the electrostatic coupling of residues in the interconnecting gap increases (e.g., Y57–X1, Y57–X2, Y57–T89, but also D85–X1, D85–X2) while the interaction between R82 and D85 is reduced. These local alterations in the region of helices B and C near the binding site are possible functionally important motions, which influence the following reaction steps. The new interaction pattern between D85, X1, X2, and T89 resembles a tightened spring, spanning the region from the retinal Schiff base to residue T89 in the lower part of the proton uptake channel. From there, the information about the changes in the active site can be transmitted over three to four additional water molecules to aspartate D96 [11,12,50].

Acid–Base Equilibria: The protonation states of the titratable residues found for the BR ground state are conserved for the ground state of the L intermediate. The changes in electrostatic interactions are converted to pK_a changes (Table III). Additional to the 2 units of stereochemical reduction, the pK_a of the retinal Schiff base is further decreased by 3.5 units. At the same time, the pK_a of D85 increases

TABLE III.   Proton distribution [(●) protonated, (O) unprotonated residue] and changes in pK_a values for key residues during the first steps in the photocycle of bacteriorhodopsin: (a) ground state; (b) reference state.

(a)	BR	L	M_1 pH < 6.2	pH > 6.2	M_2 pH < 6	6 < pH < 8.5		pH > 8.5
RSB	15.2 ●	11.7 ●	6.2 ●	6.2 O	6.0 ●	6.0 O	8.5 ●	8.5 O
D212	−3.8 O	−0.5 O	−11.3 O	−3.5 O	−5.3 O	2.6 O	−1.8 O	6.1 O
D85	−2.3 O	2.8 O	11.5 ●	17.7 ●	14.7 ●	21.1 ●	17.5 ●	23.9 ●
X1	−10.5 O	−10.4 O	−27.0 O	−19.8 O	−25.7 O	−17.5 O	−23.2 O	−15.0 O
R82	26.1 ●	28.8 ●	15.7 ●	20.0 ●	15.7 ●	18.4 ●	29.1 ●	31.8 ●
E204	20.7 ●	20.3 ●	15.5 ●	17.7 ●	6.1 ●	8.5 ●	6.1 O	8.5 O

(b)		BR	L	M_1	M_2
RSB	O	15.2	11.7	12.4	12.4
D212	O	4.5	8.1	10.2	13.7
D85	O	4.4	9.2[2.7]	17.7	21.1
X1	O	0.9	−2.4	0.7	−0.1

by 5 units, that of D212 by 3 units. The changed hydrogen bond pattern for tyrosine Y57 reduces its pK_a by 5 units. These shifts are essentially a consequence of the overall changes in the binding pocket (dominant changes in ΔpK_{back}) and not due to the shift in specific pairwise interactions.

In contrast to the titration behavior of BR, which was determined by a single, energetically well-separated neutral ground state, the L intermediate shows a more complex pattern (Fig. 5). The increased pK_a of D85 induces proton uptake at pH values ≤ 4, which leads to the admixture of a second L population (calculated 39% at pH 3) for which one of the possible proton accepting sites in the release pathway is blocked. The situation is similar to the purple–blue transition for wild type BR and confirms that bacteriorhodopsin structures with all-*trans* and 13-*cis* retinal have different pK_a's for the purple to blue transition. In the intermediate pH range from 4 to 10 proton transfer states with low activation energy ($\Delta G_{PT} \approx 1.3kT$) occur (see below) concurrently to the ground state. In the alkaline pH region (>10), the pH catalyzed release of the Schiff base proton to the aqueous medium is induced by its reduced pK_a. The drastic reduction of the pK_a for tyrosine Y57 favors its deprotonation and the formation of a salt bridge to D212. On the other hand, the salt bridge formation between R82 and D85 becomes more unfavorable compared to the BR structure. These findings can be interpreted in terms of a change in the protein heterogeneities and show that these inhomogeneities can occur also during the photocycle (photoinduced).

The titration behavior indicates the restriction of the active pH range for proton pumping to the pH region 4 to 10, as was found experimentally [25].

The L → M Transition

The experimental characterization of the L to M transition is complicated by the occurence of several kinetically distinguishable apparent M species. This prompted the proposition of different kinetic models: (A) The biphasic rise and decay may reflect separated reaction pathways of two physically different components with independent and unidirectional cycles [4–7]. (B) Besides these parallel cycle models, a single-cycle reaction sequence containing reverse reactions and two M forms is used to fit the kinetic data (e.g., [8,25]).

In this contribution, we trace possible proton transfer pathways starting from the L ground state species. The discussion of the *energetics* of the thermal proton transfer reactions in the L to M step requires the calculation of the pK_a values for donor and acceptor in the reactant and product proton configuration (see Methods). A quick overview of proton transfer facilities for different protein structures is possible through the construction of a reference proton configuration, where the protonation sites (RSB, D85, D212, and X1) are unoccupied [see Table III(b)]. This simulates the transfer of the Schiff base proton to the aqueous phase as a hypothetical intermediate step in a thermodynamic cycle going from reactant to product proton configuration. For the BR structure the smallest calculated pK_a difference for donor (RSB) and acceptor (D85) is 11 pH units. As the pK_a values approach each other in the L structure, this difference is reduced to 2.5 units. The results show also that

the transfer to D212 in the L structure is energetically less favorable (3.6 pH units). The role of D85 as catalytic proton release binding site is supported by much self-consistent experimental evidence [9,13,14]. On the other hand, mutations of the other aspartate (D212) in the active site revealed that this residue is necessary for efficient proton release [13,14]. Our results reveal its dominant contribution in adjusting the pK_a value for the retinal Schiff base. Presumably, D212 plays an essentially structural role. It may be protonated in the second half of the photocycle [10], but little could be said definitely about its involvement in proton transfer.

Our results indicate the appropriate pK_a shift for donor and acceptor residue (D85), but the structure reveals that the distance for direct proton transfer is too large (5.1 Å). Therefore, another residue must act as intermediate proton binding site. Possible transient acceptors are the tightly bond water X1 and the protein backbone peptide group (position 212, helix G) next to the Schiff base proton (Fig. 2). Our results indicate that the deprotonation of the Schiff base is initiated by the electrostatic interaction with this main chain oxygen. The measured activation energy for the L → M step in wild type, which is with 13.5 kcal/mol [8] much larger than is expected from the determined pK_a difference for donor and acceptor [8,15] of 0.6 pH units ($\triangleq \Delta G^{\#} \approx 1$ kcal/mol) points also to the involvement of an additional kinetic step.

Starting from the static L structure, the calculated activation energies for the transfer of the retinal Schiff base proton to water X1 or the peptide group are 20 kcal/mol and 18 kcal/mol, respectively. For both product states, we monitored the relaxation of chromophore and protein by means of molecular mechanics. The subsequent calculation of the acid base equilibria shows that the proton affinities for the (hypothetical M) structure with the protonated main chain peptide group are largely unchanged from those found for the L intermediate with a slightly favored transfer to aspartate D85. Even though, this state could act as "virtual intermediate" and reduce the activation barrier for following transfer reactions.

On the other hand, the protein structural reorientations which accompany the transfer of the proton from the Schiff base to water X1 change the proton affinities profoundly. The pK_a values for hydronium X1 and D85 increase drastically to a value near 18 for both groups in the reference state. The equalized affinities for X1 and D85 imply that this structure can serve as barrierless "switch state": X1 acts as a transient proton binding site and catalyzes the preparation of the release pathway for optimal unidirectional transport to D85. If the pH is larger than pK_a (RSB, 4.1) the proton is shared between X1 and D85. The relative distribution is a function of pH: For pH > 6, the proton is predominantly localized on X1 (e.g., for pH 7, we find $\langle x_{X1} \rangle = 0.61$ and $\langle x_{D85} \rangle = 0.39$), for pH 6 the average degree of protonation for both sites is 0.5, and for pH < 6 the proton is mainly found on aspartate D85 (e.g., pH 3: $\langle x_{X1} \rangle = 0.03$ and $\langle x_{D85} \rangle = 0.97$). The dissipation of 20 kcal/mol electrostatic energy (compared to L) and the increased activation energy (to 28 kcal/mol) for the back reaction ensure the irreversibility of the transfer step.

The M Intermediate(s)

During the lifetime of the M intermediates with deprotonated retinal Schiff base two kinetically different steps occur: One is a "reorientation step" [8] which guar-

antees that the reaction proceeds mainly in forward direction; the other step involves pH-dependent proton release to the extracellular and proton uptake from the cytoplasmic side of the membrane [24–26]. Both, the nature of the "reorientation step" and its location in the photocycle are disputed. One class of models [8,63] expects two M intermediates (M_1 and M_2), one with the Schiff base oriented towards the proton release pathway and one oriented towards the proton uptake pathway. Consequently, the L to M_1 step is connected to proton release, the M_2 to N step is connected to proton uptake. The other group of models [17,21,22,23] argues for reversible changes in the protein after M formation, which contribute to the restructuring of the central proton binding site. The relative sequence of protein reorientation and proton release/uptake step, however, is unclear. The analysis of the pH dependence of the proton release reaction argues for the existence of an extracellular proton release group (XH, [1,24]) whose pK_a determines the pathway for further reactions: At a pH > pK_a(XH) the photocycle proceeds with deprotonation of the release group. The L to M sequence is therefore resolved into [25] either

$$L(0) \rightleftharpoons M_1(0) \rightleftharpoons M_1(-) + H^+ \text{ (to bulk)} \rightleftharpoons M_2(-)$$

or

$$L(0) \rightleftharpoons M_1(0) \rightleftharpoons M_2(0) \rightleftharpoons M_2(-) + H^+ \text{(to bulk)}.$$
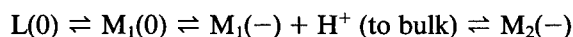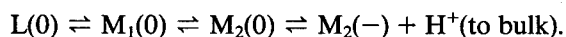
The symbols in parentheses indicate the number of protons compared to the BR ground-state. At a lower pH the cycle proceeds without deprotonation, and the proton release is delayed until after the uptake from the cytoplasmic side.

As dichroism measurements [64] give no hint for a changed orientation of the retinal throughout the lifetime of the M intermediates, the reorientation step is likely to be localized in the protein. The necessity of a protein conformational change for proton pumping is also clearly demonstrated by low temperature experiments [23]: Freezing of the protein results in a reprotonation of RSB from D85; no pumping occurs.

M_1: Our analysis of possible L → M steps reveals that the transient binding of the Schiff base proton on water X1 leads to the first stable M intermediate (named M_1; see Fig. 3) where the proton is localized on aspartate D85. We followed the accompanying reorientation of chromophore and protein by means of molecular mechanics and electrostatic calculations. Compared to the preceeding step, no further reduction in ground-state free energy occurs. The strong interaction to the backbone, which we found for the L intermediate, is relaxed. The positive charged arginine R82 moves further away from the active site towards the extracellular side of the membrane—a motion which was initiated by the positive charged hydronium in the preceeding step. Increased electrostatic interactions strengthen the connection between D85, the tightly bound waters (X1, X2) and threonine T89 in the possible proton uptake channel.

These changed electrostatic interactions lead to shifts in the pK_a values. Compared to the L or BR structure we find the situation now inverted: D85 is the high pK_a group which is always protonated, the RSB is the low pK_a residue (6.2). In order to
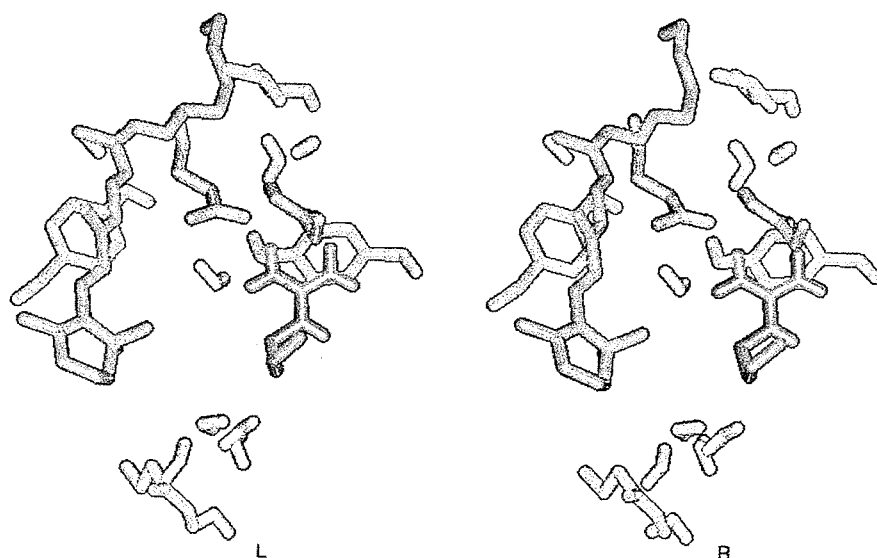
Figure 3. M_1 intermediate—stereoview of retinal and functionally important residues in
the proton release pathway.

discuss the titration behavior of the individual sites for this intermediate, we analyze
the lowest energy proton configurations as a function of pH [Table III(a)]. For the
pH region above the pK_a (RSB, 6.2) the M_1 intermediate with deprotonated RSB
will be accumulated. If the pH is below this pK_a, the charge in the protein is
determined by the uptake of a proton by the Schiff base from the aqueous medium
(Fig. 5).

We investigated also the possible involvement of one of the tightly bound waters
in the binding pocket as internal donor for the reprotonation of RSB. We find that
at least 20 kcal/mol activation energy are necessary for proton transfer from X2 to
RSB. This value is considerably lower for the M_1 structure compared to L.

The M_1–M_2 Step: In [24], the hydrogen bonded complex of R82, Y57, and a
bound water or hydronium ion is discussed as possible candidate for the proton
release (XH) group. Our results give no hint for a favorable deprotonation of one
of these in our M_1 structure. On the other hand, it was found that arginine R82 is
an essential residue which catalyzes the proton release [17]. We searched, therefore,
for a residue located near the extracellular side with a high enough pK_a to be
protonated in the initial state and whose pK_a can be influenced by the positive
charge on R82. Our candidate for the proton release group is glutamate E204! The
additional $M_1(0) \rightleftharpoons M_2(0)$ step, which precedes the proton release, is, therefore, a
reorientation of R82 from the protein interior (near the Schiff base) towards the
aqueous interface. In the structure both orientations of R82 are possible [18]. pK_a
calculations [32] and experiments with R82 mutants [15] suggest that in the initial
state the geometry that orients R82 towards the Schiff base and D85 is more likely.

A twist of the three single bonds in the arginine residue needs ca. 10 kcal/mol activation energy and brings R82 quite close to glutamate E204. As there is no direct counterion near this glutamate residue, its initial high pK_a is comparable to the one calculated for D96. A pK_a increase in the BR ground state for E204 was also found in another theoretical investigation [32]. The reorientation of the positive charge of R82 stabilizes the negative charge on E204. The motion of R82 during the photocycle was proposed earlier [10,17,54]. Large protein conformational changes during the lifetime of M were found experimentally [59]. The strong pH dependence of the M_2 electrogenicity indicates [65] that the M_1 to M_2 transition involves complex charge motions, as is expected in the proposed conformational change of the protein. The experimental $M_1 \rightarrow M_2$ rate constant (2 10^4 s^{-1} [8,25]) is in the upper limit for large conformational transitions in the protein.

M_2: Figure 4 shows the equilibrated protein structure after reorientation of arginine R82 from its location near the active site to the extracellular side of the membrane (M_2 intermediate). The most prominent change in the electrostatic interactions is a reduction of the R82–D85 interaction by 10 pH units and the concomitant increase of the R82–E204 coupling by the same amount. The interaction between the Schiff base and X1 is enlarged by 2 pH units; the strong interaction between D85, X1, X2, and T89 is maintained.

The titration behavior is determined by four energetically close-lying proton configurations, whose relative contribution is a function of the pH and accounts for the complex uptake/release pattern [Fig. 5 and Table III(a)]. The differences in
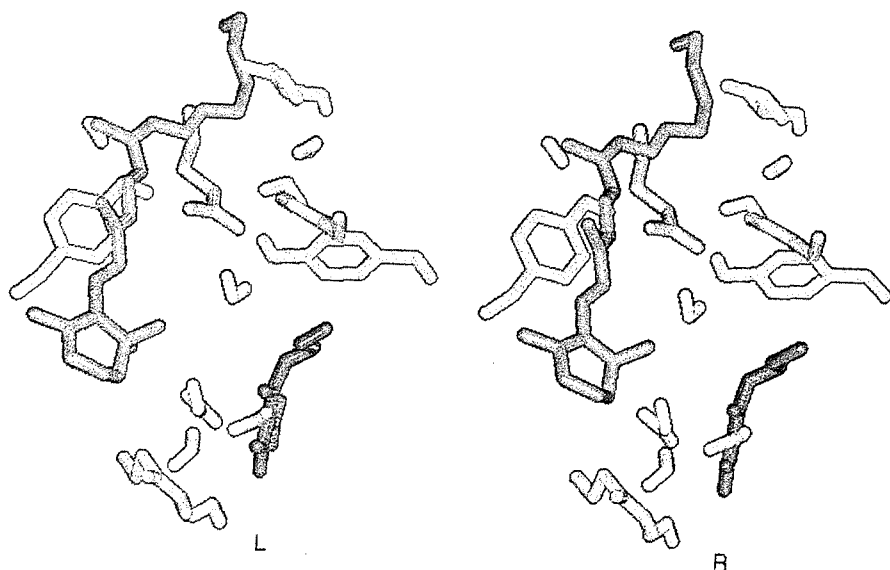
Figure 4. M_2 intermediate—stereoview of retinal and functionally important residues in the proton release pathway.
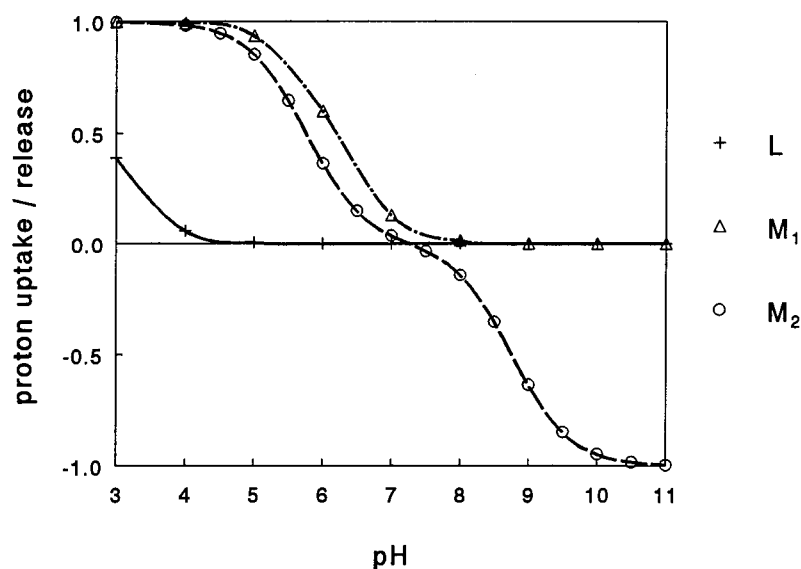
Figure 5. Calculated total charge for a Boltzmann weighted distribution of ionization states for titratable residues in the L, M_1, and M_2 intermediate (reference BR = 0).

protonation of the sites lead to multiple pK_a values for the residues. As discussed earlier, the pK_a shifts are proportional to the interaction energy values W_{ij}. For pH $<$ pK_a (RSB, 6.0), neither the Schiff base nor the glutamate E204 will deprotonate. Therefore, we find a net proton uptake compare to the L or BR structure. In an intermediate pH region ranging from pK_a(RSB, 6.0) to pK_a(E204, 8.5) two almost isoenergetic proton configurations determine the ground state. In both of them, the protein remains neutral. The proton is localized either on the Schiff base or on the glutamate E204 residue. The deprotonation of E204 rises the pK_a of the RSB by 2.5 units to pK_a(RSB) = 8.5. If the pH is increased to a value above 8.5, the Schiff base and E204 deprotonate and a net deficit of one proton in the protein occurs.

The results show that the postswitch arrangement of the protein is a prerequisite for the deprotonation of E204, the proton release group. The proton configuration with net proton uptake can be identified with the next intermediate (N) in the photocycle. Our calculations propose reprotonation of the retinal Schiff base from the aqueous medium, in contrast to the experimentally found reprotonation from a residuum located in the cytoplasmic uptake channel (e.g., D96). However, the consistent modeling of the intramolecular proton transfer in the uptake phase requires several further steps in our calculation, e.g., the explicit generation of the equilibrated protein structures for $M_2(-)$ and $M_2(+)$ as well as the improvement of the description of the solvation for the water-exposed residues.

Conclusions

Starting from the medium resolution structure of Henderson, this investigation presents a detailed model for the molecular mechanisms which lead to proton release in the first half of the bacteriorhodopsin photocycle. The essential features of the intermediates (structure, hydration, titration, spectroscopy) turn out to be consistent with experimental results. Especially, the discussion of the pairwise electrostatic interactions in the BR ground state gives insight into the individual contributions to pK_a shifts and is in agreement with mutation experiments.

Isomerization induced protein conformational changes shift the acid base equilibria of chromophore and key residues. These pK_a shifts are used as guidance for the discussion of possible proton transfer pathways. We propose the following sequence for the L to M step:

$$
\cdots L \rightarrow M_1^* \rightleftharpoons M_1 \rightleftharpoons M_2 \rightleftharpoons
\begin{cases}
\begin{array}{l}
+H^+ \\
\rightleftharpoons \quad RSBH^+, D85\text{-}H, E204\text{-}H \quad (pH < 6.0) \\[4pt]
\left\langle
\begin{array}{ll}
RSBH^+, D85\text{-}H, E204^- & \\
RSB, D85\text{-}H, E204\text{-}H & (6.0 < pH < 8.5)
\end{array}
\right. \\[8pt]
\rightleftharpoons \quad RSB, D85\text{-}H, E204^- \quad (pH > 8.5) \\
-H^+
\end{array}
\end{cases}
$$
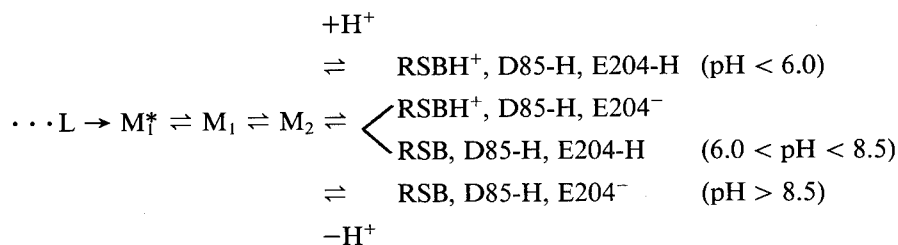
$L \rightarrow M_1^*$:   The deprotonation of the Schiff base is initiated by the electrostatic interaction to the main chain oxygen (position D212, helix G). A tightly bound water molecule (X1 in Fig. 1) acts as transient proton binding site; free energy is dissipated to the protein (irreversible switch step).

$M_1^* \rightleftharpoons M_1$:   Proton transfer from X1 to aspartate D85.

$M_1 \rightleftharpoons M_2$:   Arginine R82 moves towards the extracellular side and induces a pK_a reduction of the proton release group, identified as glutamate E204 (protein reorientation step).

$M_2 \rightleftharpoons \cdots$ pH-dependent proton release and uptake.

Besides this linear sequence including back reactions, the existence of parallel photocycles starting from different BR species cannot be excluded. Our results point to pH-dependent charge heterogeneities, which vary during the photocycle.

Acknowledgments

Bibliography

[1] J. K. Lanyi, Experientia **49**, 514–517 (1993).
[2] R. A. Mathies, S. W. Lin, J. B. Ames, and W. T. Pollard, Annu. Rev. Biophys. Biophys. Chem. **20**, 491–518 (1991).

[3] R. R. Birge, Biochim. Biophys. Acta **1016**, 293–327 (1990).

[4] R. Diller and M. Stockburger, Biochemistry **27**, 7641–7651 (1988).

[5] W. Eisfeld, C. Pusch, R. Diller, R. Lohrmann, and M. Stockburger, Biochemistry **32**, 7196–7215 (1993).

[6] J. H. Hanamoto, P. Dupuis, and M. A. El-Sayed, Proc. Natl. Acad. Sci. USA **81**, 7083–7087 (1984).

[7] M. A. El-Sayed, Acc. Chem. Res. **25**, 279–286 (1992).

[8] G. Váró and J. K. Lanyi, Biochemistry **30**, 5016–5022 (1991).

[9] B. Heβling, G. Souvignier, and K. Gerwert, Biophys. J. **65**, 1929–1941 (1993).

[10] M. S. Braiman, R. Mogi, T. Marti, L. J. Stern, F. G. Khorana, and K. J. Rothschild, Biochemistry **27**, 8516–8520 (1988).

[11] K. Gerwert, B. Hess, J. Soppa, and D. Oesterhelt, Proc. Natl. Acad. Sci. USA **86**, 4943–4947 (1989).

[12] H. Otto, T. Marti, M. Holz, T. Mogi, M. Lindau, H. G. Khorana, and M. P. Heyn, Proc. Natl. Acad. Sci. USA **86**, 9228–9232 (1989).

[13] H. Otto, T. Marti, M. Holz, T. Mogi, L. J. Stern, F. Engel, H. G. Khorana, and M. P. Heyn, Proc. Natl. Acad. Sci. USA **87**, 1018–1022 (1990).

[14] R. Needleman, M. Chang, B. Ni, G. Váró, J. Fornés, S. H. White, and J. K. Lanyi, J. Biol. Chem. **266**, 11,478–11,484 (1991).

[15] L. S. Brown, L. Bonet, R. Needleman, and J. K. Lanyi, Biophys. J. **65**, 124–130 (1993).

[16] Y. Cao, G. Váró, A. L. Klinger, D. M. Czajkowsky, M. S. Braiman, R. Needleman, and J. K. Lanyi, Biochemistry **32**, 1981–1990 (1993).

[17] S. P. Balashov, R. Govindjee, M. Kono, E. Imasheva, E. Lukashev, T. G. Ebrey, R. K. Crouch, D. R. Menick, and Y. Feng, Biochemistry **32**, 10,331–10,343 (1993).

[18] R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing, J. Mol. Biol. **213**, 899–929 (1990).

[19] P. Hildebrandt and M. Stockburger, Biochemistry **23**, 5539–5548 (1984).

[20] H. J. M. DeGroot, G. S. Harbison, J. Herzfeld, and R. G. Griffin, Biochemistry **28**, 2245–3353 (1989).

[21] S. P. A. Fodor, J. B. Ames, R. Gebhard, E. M. M. van der Berg, W. Stoeckenius, J. Lugtenburg, and R. A. Mathies, Biochemistry **27**, 7097–7101 (1988).

[22] M. S. Braiman, O. Bousché, and K. J. Rothschild, Proc. Natl. Acad. Sci. USA **88**, 2388–2392 (1991).

[23] P. Ormos, K. Chu, and J. Mourant, Biochemistry **31**, 6933–6937 (1992).

[24] L. Zimányi, G. Váró, M. Chang, B. Ni, R. Needleman, and J. K. Lanyi, Biochemistry **31**, 8535–8543 (1992).

[25] Y. Cao, L. S. Brown, R. Needleman, and J. K. Lanyi, Biochemistry **32**, 10,329–10,248 (1993).

[26] L. Zimányi, Y. Cao, R. Needleman, M. Ottolenghi, and J. K. Lanyi, Biochemistry **32**, 7669–7678 (1993).

[27] A. Maeda, J. Sasaki, Y. Shichida, R. Yoshizawa, M. Chang, B. Ni, R. Needleman, and J. K. Lanyi, Biochemistry **31**, 4684–4690 (1992).

[28] A. Warshel, Methods Enzymol. **127**, 578–587 (1986).

[29] A. Warshel, Photochem. Photobiol. **30**, 285–290 (1979).

[30] D. Bashford and M. Karplus, J. Phys. Chem. **95**, 9556–9561 (1991).

[31] A.-S. Yang, M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig, Proteins Struct. Funct. Genet. **15**, 252–265 (1993).

[32] D. Bashford and K. Gerwert, J. Mol. Biol. **223**, 473–486 (1992).

[33] K. A. Sharp and B. Honig, Annu. Rev. Biophys. Biophys. Chem. **19**, 301–332 (1990).

[34] C. Scharnagl, C. Cometta-Morini, and S. F. Fischer, Int. J. Quantum Chem. Quantum Biol. Symp. **20**, 199–212 (1993).

[35] C. Cometta-Morini, C. Scharnagl, and S. F. Fischer, Int. J. Quantum Chem. Quantum Biol. Symp. **20**, 89–106 (1993).

[36] R. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **4**, 187–199 (1983).

[37] M. J. S. Stewart, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902–3909 (1985).

[38] R. Henderson and P. N. T. Unwin, Nature **257**, 28–32 (1975).
[39] L. Stryer, *Biochemistry* (Freeman, New York, 1988).
[40] R. Stewart, *The Proton: Application to Organic Chemistry* (Academic, New York, 1985).
[41] P. Tavan, K. Schulten, and D. Oesterhelt, Biophys. J. **47**, 415–430 (1985).
[42] M. Nonella, A. Windemuth, and K. Schulten, Photochem. Photobiol. **54**, 937–948 (1991).
[43] F. Zhou, A. Windemuth, and K. Schulten, Biochemistry **32**, 2291–2306 (1993).
[44] W. J. Joergensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. **79**, 926–935 (1983).
[45] G. Dueck and T. Scheuer, J. Comp. Phys. **90**, 161–175 (1990).
[46] A. Warshel, Z. T. Chu, and J.-K. Hwang, Chem. Phys. **158**, 303–314 (1991).
[47] P. O. J. Scherer and S. F. Fischer, in *Chlorophylls*, H. Scheer, Ed. (CRC Press, Boca Raton, FL, 1991).
[48] A. Fahmy, F. Siebert, M. Grossjean, and P. Tavan, J. Mol. Struct. **214**, 257–288 (1989).
[49] M. S. Braiman, P. L. Ahl, and K. J. Rothschild, Proc. Natl. Acad. Sci. USA **84**, 5221–5225 (1987).
[50] Y. Cao, G. Váró, M. Chang, B. Ni, R. Needleman, and J. K. Lanyi, Biochemistry **30**, 10,972–10,979 (1991).
[51] A. Maeda, J. Sasaki, Y. Shichida, and R. Yoshizawa, Biochemistry **31**, 462–467 (1992).
[52] G. Papadopoulos, N. A. Dencher, G. Zaccai, and G. Büldt, J. Mol. Biol. **214**, 15–19 (1990).
[53] M. R. Gunner and B. Honig, in *The Photosynthetic Bacterial Reaction Centers II*, J. Breton and A. Verméglio, Eds. (Plenum, New York, 1992), pp. 403–410.
[54] K. J. Rothschild, Y.-W. He, S. Sonar, T. Marti, and H. G. Khorana, J. Biol. Chem. **267**, 1615–1622 (1992).
[55] S. Druckman, M. Ottolenghi, A. Pande, J. Pande, and R. H. Callendar, Biochemistry **21**, 4953–4959 (1982).
[56] U. Fischer and D. Oesterhelt, Biophys. J. **28**, 211–230 (1979).
[57] J. K. Lanyi, J. Tittor, G. Váró, G. Krippahl, and D. Oesterhelt, Biochim. Biophys. Acta **1099**, 102–110 (1992).
[58] L. S. Brown, L. Zimányi, R. Needleman, M. Ottolenghi, and J. K. Lanyi, Biochemistry **32**, 7679–7685 (1993).
[59] K. Fukuda and T. Kouyama, Biochemistry **31**, 11,740–11,747 (1992).
[60] S. P. Balashov, R. Govindjee, and T. G. Ebrey, Biophys. J. **60**, 475–490 (1991).
[61] R. R. Birge and T. M. Cooper, Biophys. J. **42**, 61–69 (1983).
[62] M. Rohr, W. Gärtner, G. Schweitzer, A. R. Holzwarth, and S. E. Braslavsky, J. Phys. Chem. **96**, 6055–6061 (1992).
[63] P. Tavan and K. Schulten, Biophys. J. **50**, 81–89 (1986).
[64] H. Otto and M. P. Heyn, FEBS Lett. **293**, 111–114 (1991).
[65] C. Gergely, C. Ganea, G. Groma, and G. Váró, Biophys. J. **65**, 2478–2483 (1993).

Conformational Analyses of the Polymorphism of Triglycerides

CHRISTINE CULOT AND FRANCOIS DURANT

*Laboratoire de Chimie Moléculaire Structurale, Facultés Universitaires Notre-Dame de la Paix,
61 rue de Bruxelles, B-5000 Namur, Belgium*

DAVID H. MOSLEY AND JEAN MARIE ANDRE

*Laboratoire de Chimie Théorique Appliquée, Facultés Universitaires Notre-Dame de la Paix,
61 rue de Bruxelles, B-5000 Namur, Belgium*

DANIEL P. VERCAUTEREN

*Laboratoire de Physico-Chimie Informatique, Facultés Universitaires Notre-Dame de la Paix,
61 rue de Bruxelles, B-5000 Namur, Belgium*

Abstract

Triglycerides, the major components of neutral lipids, are important biomaterials, as they take part in the edification of membranes. In this perspective, the consideration of biological membranes at a molecular level requires detailed knowledge of the preferred conformations of the triglycerides in their various polymorphic forms. In this context, we adapted a molecular modeling approach, which allows the simulation the three-dimensional structure of the different polymorphic forms (α, β', and β) valid for any triglyceride. Their conformational analysis is based on molecular mechanics calculations, as follows: First, a large number of isolated molecular structures were generated in a systematic structure-tree analysis. For their generation, atomic charges within the Mulliken scheme, calculated at the *ab initio* RHF–LCAO–MO–SCF level (6-31G), were considered. The lowest-energy conformers were, next, correlated with experimental data (NMR, powder X-ray diffraction) in order to select α, β', or β structures. Then, in a second step, these selected conformers were assembled, in head-to-tail dimers in order to form a monolayer. For this step, the use of potential-derived atomic charges is known to be more suitable. In this study, we consider triglycerides derived from the predominant fatty acids, i.e., stearic, elaidic, and oleic acids. © 1994 John Wiley & Sons, Inc.

Introduction

Triglycerides, triesters of glycerol and fatty acids, are important raw materials. Being the main components of natural fats, they are widely used in the food industry [1]. However, they also fulfill important biochemical functions (edification of membranes, transport of fats, etc.) [2–5]. The nature of the fatty acids, saturated or not, as well as their position on the glycerol backbone, generates a large number of triglyceridic structures: monoacid, diacid, asymmetric, symmetric, mixed saturated and unsaturated, etc. Furthermore, triglycerides may exist in various polymorphic forms, α, β', or β, according to different lateral crystal packing. The α form structure is similar to the melted state, where hydrocarbon chains are in free rotation and presents a tuning-fork conformation [Fig. 1(a)]. The β' and β forms

CULOT ET AL.

a) α   b) β' , β
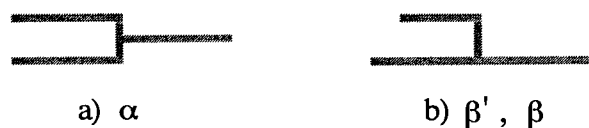
Figure 1.   Schematization of the polymorphic (a) α form: tuning-fork and (b) β' and β
forms: chair.

present a chair structure [Fig. 1(b)], in which the chains are tilted. The presence
of different types of unsaturation influences the existence and the stability of the
polymorphic forms [6,7]. Moreover, triglycerides may also pack in different lon-
gitudinal arrangements, i.e., in double (L-2) or triple (L-3) hydrocarbon chain
lengths (Fig. 2). For example, monoacid triglycerides crystallize in a double hy-
drocarbon chain length, whereas *cis* unsaturated diacid compounds pack in a triple
hydrocarbon chain length.

The diversity of such molecules as well as their associated polymorphism confer
to these lipids complex structural behavior, which is related to particular physi-
co-chemical properties such as fluidity or biochemical activity [8]. It is thus essential
to have a thorough knowledge of the structure and molecular organization of the
different polymorphic forms in order to understand or even predict some particular
membrane characteristics. Unfortunately, little structural information, coming from
monocrystal X-ray diffraction studies, is available [9-12], and, moreover, the
method can be applied only to stable polymorphic forms.

For these reasons, we wished to develop a molecular modeling approach, to
simulate the three-dimensional structure of the α, β', or β crystalline forms of any
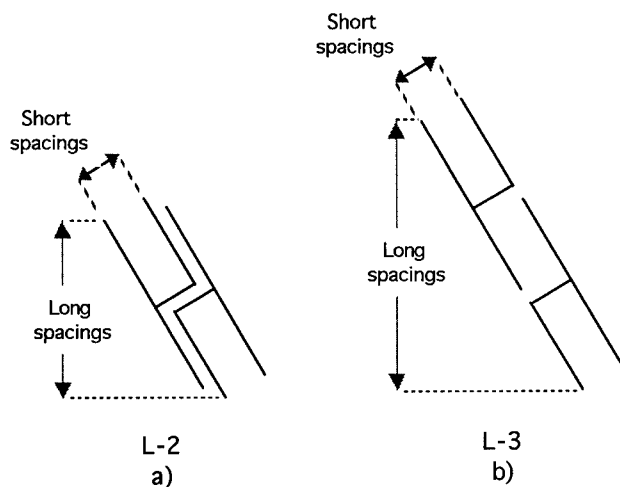
Figure 2.   Schematization of the longitudinal packing according to (a) a double L-2 and
(b) a triple L-3 hydrocarbon chain length.

triglyceride. Our attention was, therefore, directed toward a molecular mechanics approach already described in the literature and successfully applied to phospholipid membranes, lipid helices, drugs, and peptides [13].

In this study, we focus on the particular structural behavior of monoacid triglycerides derived from stearic, elaidic, and oleic acids (the most abundant fatty acids), i.e., tristearin (which is saturated), trielaidin (*trans* unsaturated), and triolein (*cis* unsaturated) (Fig. 3).

The article is organized as follows: In the first part, the strategy concerning the conformational analysis of isolated conformers as well as their assembly is described, by considering trilaurin, a C_{12} triglyceride whose three-dimensional structure is available [10]. In this step, particular focus is put on the calculation of the atomic charges, needed in the electrostatic contribution.

In a second part, the procedure defined is applied to tristearin, trielaidin, and the triolein model: three C_{18} triglycerides. As for trilaurin, the three-dimensional structure of isolated compounds is analyzed first. Particular conformations are then retained, by correlation with experimental data (^{13}C-MAS-CP-NMR, powder X-ray diffraction). These conformers are then assembled, in head-to-tail bimolecular conformations, whose stability is also correlated with experimental values (differential scanning calorimetry). Finally, the lowest-energy dimers will be associated to form a monolayer, approaching, thereby, the three-dimensional packing of the different polymorphic forms.

CH_2-O-CO-R   (I)
CH -O-CO-R   (II)
CH_2-O-CO-R   (III)

			Radical (R)
Trilaurin	C_{12}		-$(CH_2)_{10}$-CH_3
Tristearin	C_{18}		-$(CH_2)_{16}$-CH_3
Triolein	C_{18}	one cis double bond	-$(CH_2)_7$-CH=CH-$(CH_2)_7$-CH_3
Trielaidin	C_{18}	one trans double bond	-$(CH_2)_7$-CH=CH-$(CH_2)_7$-CH_3

Reduced model

CH_2-O-CO-$(CH_2)_2$-CH_3 (I)
CH -O-CO-$(CH_2)_4$-CH_3 (II)
CH_2-O-CO-$(CH_2)_2$-CH_3 (III)
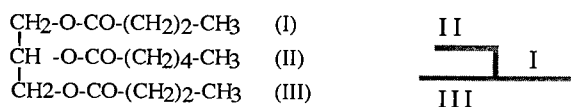
Figure 3.   Structure formulas of the considered triglycerides.

Methods and Strategy

The conformational analysis method chosen, widely applied to lipid and phospholipid membranes, is based on molecular mechanics calculations and involves a three-step procedure [13]: first, the generation of the three-dimensional structure of an isolated molecule; second, the assembly of two molecules to form a dimer; and third, the formation of a monolayer, by association of dimers.

Conformational Analysis of the Isolated Molecule

For the generation of the isolated models to be used later for the assemblies, the procedure is in two steps: First, a structure-tree analysis to generate a certain number of conformers of high probability (low energy) and, second, an energy minimization of the obtained high probability conformers. For both parts, one needs thus to use a potential function. The total conformational energy is considered as a sum of three terms:

1. A van der Waals energy contribution defined by a Buckingham's pairwise function between all pairs of nonbonded atoms i and j:

$$E_{vdW} = \sum_{ij} A_{ij} e^{-B_{ij} r_{ij}} - \frac{C_{ij}}{r_{ij}^6}, \tag{1}$$

where r_{ij} is the distance between all atoms, and A_{ij}, B_{ij}, and C_{ij} are coefficients assigned to atom pairs. The values chosen for the constants are those proposed by Liquori and co-workers [14,15].

2. An electrostatic energy contribution described by a Coulombic interaction term between atomic point charges:

$$E_{elec} = 332 \sum_{ij} \frac{q_i q_j}{\varepsilon_{ij} r_{ij}}, \tag{2}$$

where ε_{ij} is the dielectric constant, and q_i and q_j, the atomic charges calculated as described later. To simulate biomembranes, mainly constituted by phospholipids of 16 or 18 carbon atoms (mean length of 20 Å), one has to consider, depending on the distance between each pair of atoms, a linear variation of ε_{ij} from 1 to 16 ($\varepsilon = 16$ is the value for a phosphate group) up to a distance of 20 Å. E_{elec} is given in kcal/mol when r_{ij} is expressed in Å and q_i and q_j in electron charge units (332 is a conversion factor, allowing the expression of the energy in kcal/mol).

3. A torsional energy contribution defined as

$$E_{tors} = \sum_{ij} \frac{U_{ij}}{2} (1 + \cos \Phi_{ij}), \tag{3}$$

where Φ_{ij} corresponds to the various torsional angles around the C—C and C—O single bonds, and U_{ij}, to the energy barrier between eclipsed and staggered

conformations. U_{ij} is set to 2.8 kcal/mol for a C—C bond and 1.8 kcal/mol for a C—O bond [16].

Starting from an all-*trans* geometry, and considering standard values for the interatomic distances and bond angles [16], a large number of conformers is generated by a structure-tree analysis [17]. If systematic 60° changes are applied for n selected torsional angles, 6^n conformers are generated. For each conformer, the total energy is computed as well as the statistical weight (according to the Maxwell–Boltzmann equation).

The sheets of the trees, i.e., the structures having a statistical weight above 1%, are then submitted to a simplex minimization procedure, with a precision of 5° for each torsional angle [18]. At the end of the minimization, the probability of existence is associated to each of the three-dimensional conformers generated.

Assembly of Monolayers

For the assembly of dimers, as well as that for monolayers, the total energy is calculated as the sum of three terms: a van der Waals contribution (E_{vdW}) and the electrostatic energy (E_{elec}) as defined above, plus a transfer energy contribution (E_{tr}). No torsional term is considered here as the models will be kept rigid. The transfer energy term is defined for a molecule as the sum of all transfer energy changes associated with the transfer of each individual atom from the hydrophilic to the hydrophobic phase [13]. These values of transfer energy per atom have been derived from total transfer energies compiled for a series of chemical analogs [19] and are equal to −1.5 kcal/mol for a C_{sp^2} atom, −2.4 kcal/mol for a C_{sp^3} atom, 1.0 kcal/mol for H, and 2.8 kcal/mol for O [20].

In the first step, to mimic the crystalline structure in which molecules are oriented in opposite directions, the conformers are assembled in a head-to-tail bimolecular configuration. One molecule, previously rotated from 180°, is moved toward another, taking into account steric constraints. The system is then submitted to an interaction energy minimization procedure, by translation (step of 0.5 Å) and rotation (step of 2.5°) along and about the three orthogonal axes. Starting from different bimolecular approaches, this sequential assembly can lead to different dimer geometries, all located within a range of energies of only ±5 kcal/mol. In our case, only the lowest-energy dimers are retained. In a second step, using the same assembly procedure, dimers are brought toward each other and their interaction energy is minimized. The procedure is then repeated until a monolayer of 16 molecules (8 dimers) is formed.

The generation of both the isolated forms and monolayers have been performed using the PC-Molecular and PC-TAMMO+ (Theoretical Analysis of Membrane Molecular Organization) procedures [13], on a Olivetti CP486 microcomputer equipped with an Intel 80486 processor. Graphics have been drawn with the PC-MGM+ (Molecular Graphics Manipulation) program [13].

Calculation of Atomic Charges

If it is usually accepted that the bonding terms are satisfactorily well described by various energy functions (such as the one proposed for the van der Waals or

dihedral contribution) for biological macromolecules, there is, however, more un-certainty regarding the electrostatic contributions. Due to their slow variation with distance, the best possible treatment of the Coulomb term is essential, especially for the evaluation of conformational equilibria and thermodynamics properties. In our case, as the electrostatic term may contribute to a large part of the total energy, a most precise evaluation of the atomic charges is desirable. More precisely, it has already been shown in a previous work dedicated to the comparison of atomic charges to be integrated into conformational analyses of neutral lipids that charges generated by semiempirical methods such as AM1 were not suitable [21]. Hence, the charges for our model have been calculated at the *ab initio* RHF–LCAO–MO–SCF method [22], with the GAUSSIAN 92 series of program [23]. Using the 6-31G basis set, the atomic charges were obtained both from the Mulliken approach [24] and also derived from the molecular electrostatic potential, proposed by Kollman [25,26]. Their respective application as well as the choice of the basis set is detailed in the discussion.

All *ab initio* MO computations were performed on a cluster of IBM RISC 6000 Model 560 computers of the Scientific Computing Facility Center (Namur-SCF) of the University of Namur.

Results and Discussion

Atomic Point Charge Evaluation

First, several test calculations were performed on reduced models of trilaurin (i.e., the glycerol moiety with 3 and 5 methylene groups for chain II and chains I and III, respectively [Fig. 3]) considered in its crystalline state [9]. The hydrogens were positioned with standard distances, bond angles, and torsional angles depending on the carrier atom. The atomic charges were derived from ca. 2500 molecular electrostatic potential values considered around the molecule with three different basis sets: 6-31G, 6-31G*, and 6-31G** (* = +d polarization functions on heavy atoms and ** = +d functions on heavy atoms and p functions on hydrogens, resulting in 280, 424, and 520 basis functions, respectively). The maximum dif-ferences between the charges obtained with the two basis including polarization functions is very small, only ±0.04 e^-, whereas the largest difference between the values obtained with 6-31G and 6-31G* can reach ±0.2 e^- in the case of the carbonyl atoms. Bearing those differences in mind, the 6-31G basis set was chosen for the further computations as a cost-quality compromise. Within the 6-31G basis set, the number of basis functions is 535, 775, 775, and 787 for trilaurin, trielaidin, triolein, and tristearin, respectively. The calculations were done using the direct SCF approach with the standard bielectron integral and convergence thresholds for single-point calculations.

For the structure-tree generation, since the torsional angles of the molecule are continuously varying, the use of point-derived charges is inadequate as they are extremely dependent upon the overall conformation. Hence, charges obtained with the Mulliken population analysis were taken into account in the electrostatic con-tribution when generating the three-dimensional structure of the isolated model.

More particularly, one considered the Mulliken atomic charge values of the glycerol region and the terminal carbons for the trilaurin and tristearin models and, when needed, appropriate charges around the double bonds for trielaidin and triolein (Fig. 4).

However, for the assembly of dimers and monolayers, since the total conformation of the molecule is then kept rigid (only translation and rotation are taken into account), the use of charges derived from the surrounding potential, and thus more sensitive to the global molecular conformation, can be considered. It is important to stress that their use has been recommended when considering intermolecular interactions [25,26]. Hence, potential-derived charges were calculated for each of the conformers that will be selected for the assembly.
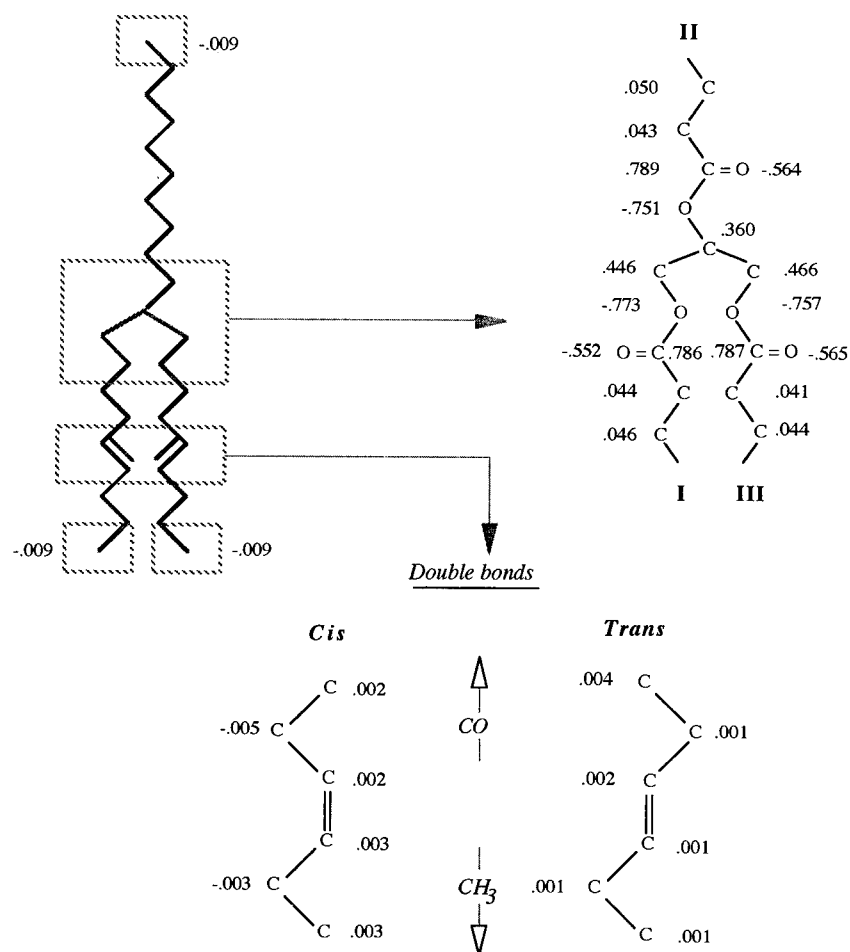
Figure 4. Atomic charges (e^-) considered for the conformational analysis of the isolated trilaurin model (6-31G basis set, Mulliken population analysis).

Conformation Analysis Strategy

To adapt the procedure to neutral lipids, a detailed study of the structure of trilaurin, a monoacid C_{12} triglyceride for which the crystalline structure is well known, was undertaken [7]. Besides the test calculations regarding the atomic charge values (basis set and method), first, one considered, in the minimization step, different values of the dielectric constant (as mentioned earlier, ε was allowed to vary in the structure-tree generation part). With increasing discrete values of ε (1, 3, 16, and 80), the contribution of the electrostatic term logically decreased but the statistical repartition of the different conformers remained unchanged. Thus, a value of $\varepsilon = 1$ was chosen for the further calculations. Second, three different structure-tree strategies, involving torsional twists indicated along the three [tree no. 1, Fig. 5(a)] or two [trees nos. 2 and 3, Fig. 5(b)] hydrocarbon chains were tested. For structure-tree no. 2, twists around chain II were not considered in order to keep the chain in the opposite direction to chains I and III. For structure-tree no. 3, twists around chain III were not allowed to force the parallelism between chains I and III. The most stable conformers generated by structure-tree no. 1 are mostly extended conformations [Fig. 6(a)] and, in a minor portion, folded conformations [Fig. 6(b)]. The most stable conformers generated by structure-trees nos. 2 and 3 are mostly folded and the extended ones present a tuning-fork conformation, representative of the α form only. As only extended conformations are able to fit with the experimentally observed α, β', or β form, the structure-tree analysis considering variations of the torsional angles along the three chains is preferred. The search method has been further optimized in two levels to avoid an excessive time-consuming procedure. The variation of the first eight torsional angles
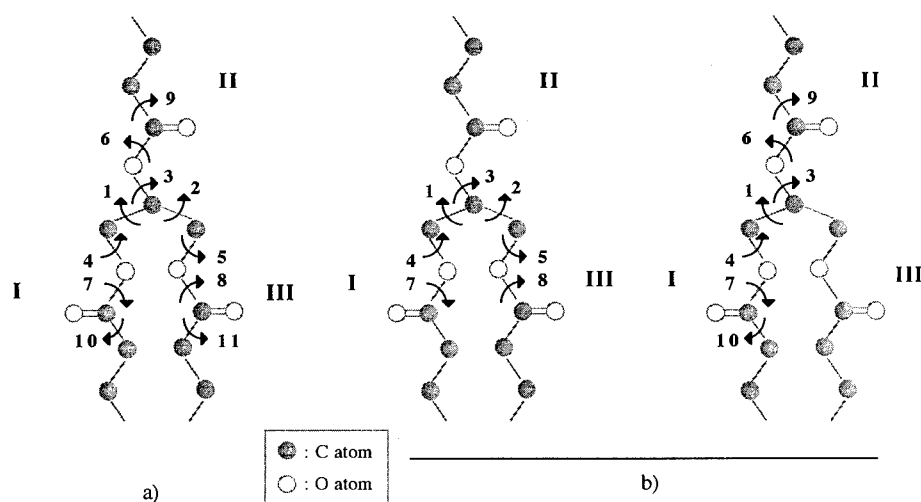
Figure 5.   Torsional angles along (a) three and (b) two hydrocarbon chains, considered for the different structure-tree analyses.
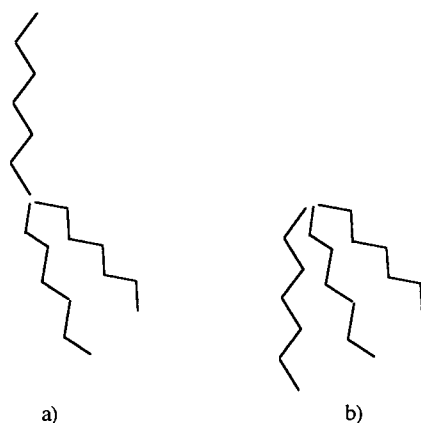
Figure 6. Classification of the modelized conformers into (a) extended structure and (b) folded structure.

[1–8, Fig. 5(a)] generates 6^8 conformers, with those having a statistical weight above 3% being kept. A second variation of angles 6–11 generates 6^6 conformers, and those having a probability of existence larger than 1% retained. The retained models need then to be classified and particular ones have to be selected to represent particular polymorphic forms. Therefore, the most probable modelized conformers are next divided in extended [Fig. 6(a)] and folded [Fig. 6(b)] structures, which are finally correlated with experimental data.

More precisely, hypotheses relative to the three-dimensional structure can be deduced by ^{13}C high resolution solid-state NMR (MAS-CP) and, more particularly, by the analysis of the chemical shifts for the carbonyl and the glyceridic carbons [27]:

- For the α form, the spectra show two peaks (for chains I and III, and chain II, respectively), possessing a symmetrical environment or a tuning-fork structure [Fig. 1(a)].
- For the β' and β forms, one observes three peaks: the β' and β forms presenting an asymmetrical or chair structure [Fig. 1(b)]. According to the NMR data for trilaurin, the extended conformers obtained can thus be classified in α form [Fig. 7(a)] or β' and β forms [Fig. 7(b)].

From powder X-ray diffraction, measurements of the distances between chain-end carbons, i.e., the long spacings (Fig. 2), can be determined, for each triglyceride under their various polymorphic forms (Table I). Hence, by comparison of the distances between the chain ends for the different conformers and the values of long spacings for each polymorphic form, particular conformers that are able to mimic α, β', and β forms were selected.

The selected conformers can then be assembled into monolayers. To approximate crystalline lattices, in which triglycerides are packed in opposite directions, two
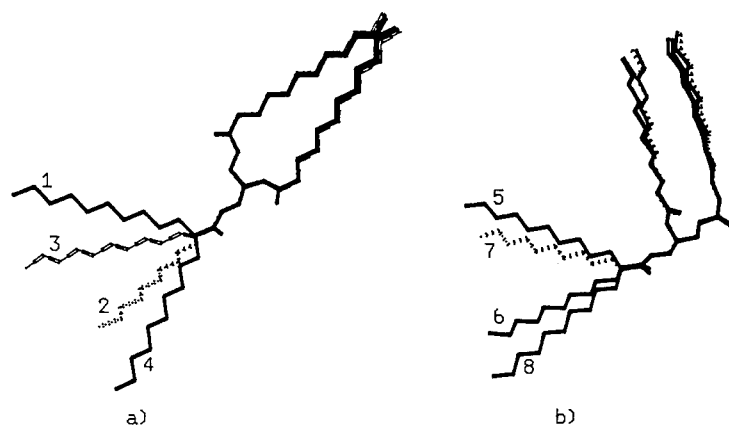
Figure 7.   Classification of the eight more probable modelized conformers of trilaurin in
(a) α-forms (nos. 1–4) and (b) β' or β forms (nos. 5–8).

molecules are first assembled sequentially in a head-to-tail bimolecular confor-
mation, taking into account steric constraints, and their interaction energy mini-
mized as described earlier. The lowest-energy dimers are then repeated to form a
monolayer.

The whole conformational analysis procedure proposed is summarized in Fig-
ure 8.

Monoacid Triglyceride Isolated Conformers

The procedure, illustrated in Figure 8, has also been applied to C_{18} monoacid
triglycerides, i.e., tristearin, trielaidin, and triolein. The results of the structure-tree
analyses are presented in Figure 9. At the first level (variation of the angles 1–8),
conformers having a statistical weight larger than 3% were kept: six for tristearin,
five for trielaidin, and seven for triolein. Starting from these conformations, one
obtains at the level 2 (variation of the angles 6–11) 12, 14, and 12 low-energy
structures for tristearin, trielaidin, and triolein, respectively. As explained earlier,
these conformers are then minimized. The most important are presented in Figure
10. All probable conformers for triolein have a folded structure (Fig. 10), evidently

TABLE I.   Values of the long spacings (Å) observed for the
polymorphic forms of trilaurin, tristearin, and trielaidin.

	α Form	β' Form	β Form
Trilaurin	34	31	30
Tristearin	53	51	48
Trielaidin	51		45

Isolated molecule
(Mulliken atomic charges)

$$E = E_{VDW} + E_{Cb} + E_{tors}$$

Initial Structure

Structure tree analysis
n°1 (angles 1 —8 / 6 - 11)

Modelized Conformers

Analysis of conformers
→ Structures

Extended   Folded

Correlations with experimental
data

- NMR

- XRD :   long spacings

α, β', β Forms

Assembly
(Potential derived atomic charges)

$$E = E_{VDW} + E_{Cb} + E_{tr}$$

"Manual"

Formation of "dimers"

Association in monolayer

Figure 8.   Diagram of the conformational analysis procedure of triglycerides.

caused by the presence of a *cis* double bond in all three chains. This structure has
been shown to exist in oil–water/air–water interfaces or inside membranar phos-
pholipids [28], but does not resemble any crystalline form reported so far. Therefore,
as for the saturated trilaurin, one will focus on the tristearin and trielaidin models
that produce extended structures comparable to the crystalline forms observed ex-

　　　　　　　　　　CULOT ET AL.
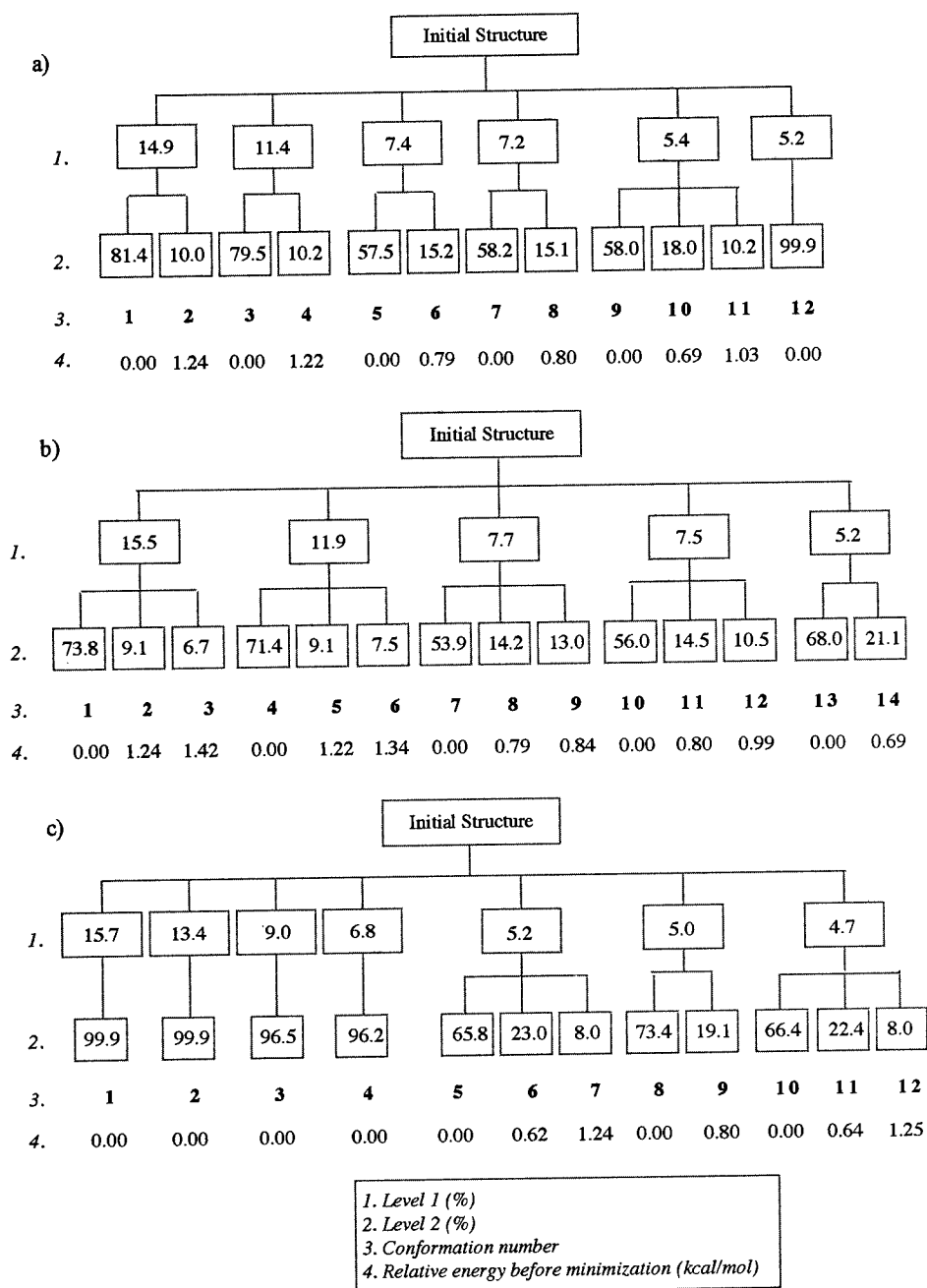
Figure 9.　Structure-trees of (a) tristearin, (b) trielaidin, and (c) triolein.
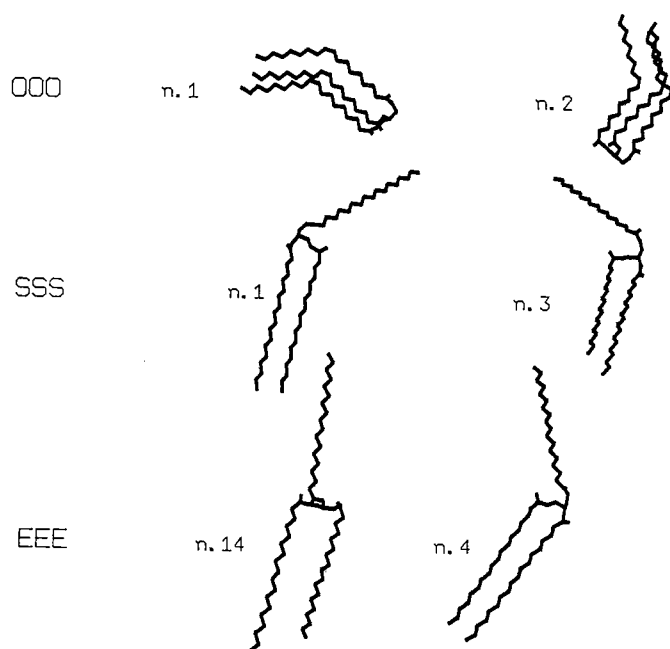
Figure 10. Structures of the most significant conformers obtained for triolein (OOO), tristearin (SSS), and trielaidin (EEE).

perimentally. Those extended models were then classified (Table II) in order to select conformers that can mimic the various polymorphic forms (Fig. 10).

For tristearin, the structures retained were

- no. 1, a tuning-fork structure, to represent the α form, and
- no. 3, a chair structure, to represent the β' or β forms.

For trielaidin, the chosen structures were

- no. 4, a chair structure, which can mimic a β form, and
- no. 14, a tuning-fork structure, which can mimic an α form.

TABLE II. Classification of the conformers obtained for tristearin and trielaidin.

	Folded structure	Extended structure	
		Chair	Tuning-fork
Tristearin	6, 8, 9, 11, 12	3, 4,	1, 2, 5, 7, 10
Trielaidin	3, 8, 11, 13	1, 2, 4, 5, 6	7, 9, 10, 12, 14

The Cartesian coordinates of the heavy atoms for the selected conformers of tristearin and trielaidin are presented in Tables III and IV, respectively.

Bimolecular and Monolayer Assemblies

The head-to-tail bimolecular assemblies are presented in Figure 11. For each type of assembly, the intra- and intermolecular distances between each pair of atoms of the three parallel chains have been determined and compared to the values of short spacings (Fig. 2) measured experimentally by powder X-ray diffraction for the different polymorphic forms. The mean distances calculated for conformers no. 3 for tristearin and no. 4 for trielaidin are, within ± 0.2 Å, close to the values of short spacings characteristics of a β form, i.e., 4.6, 3.9, and 3.6 Å. The distances calculated for conformer no. 1 of tristearin vary around 4.2 Å, which is the value of short spacings observed for an α form.

Here, also, correlations have been made with experimental data, coming from differential scanning calorimetry (DSC). By this technique, and more precisely by the measurement of the enthalpy of fusion (ΔH) of the α, β', and β forms, one can evaluate the stability of the different polymorphic forms. To obtain a corresponding theoretical relative stability value for each dimer, we calculated the mean interaction energy between two molecules (E_{int}), as the sum of the van der Waals, electrostatic, and transfer energies.

From Table V, one can see that the interaction energy values, calculated for the dimers, are well correlated with those of ΔH, i.e.:

- For saturated compounds, as the hydrocarbon chain length increases, the stability of the lattice increases as well; the values of E_{int} are much lower for tristearin than for trilaurin.
- The β form is more stable than is the α form.
- The presence of unsaturations, in the case of trielaidin, reduces the stability of both the α and β forms.

Theoretical structure results obtained with our molecular modeling approach thus corroborate nicely with the available experimental data.

Finally, the most probable dimers obtained for trilaurin, tristearin, and trielaidin have been assembled to form a monolayer in order to simulate the three-dimensional packing of α and β forms in a double hydrocarbon chain length. For illustration, a monolayer of trielaidin, under the α and β forms, has been presented in Figure 12.

Conclusions

The resolution of the three-dimensional structure of triglycerides by monocrystal X-ray diffraction is rather complicated, essentially due to the difficulty in obtaining monocrystals. To our knowledge, only three crystalline structures have been solved to date [9–12]. Our laboratory experiments, despite the development of a particular crystallization system, and after a large number of crystallizations of several different compounds, finally enabled us to obtain one monocrystal of β-trielaidin (twins

TABLE III. Cartesian coordinates of the heavy atoms for the selected conformers of tristearin.

| | No. 1 (α form) | | | No. 3 (β form) | | |
	x	y	z	x	y	z
C	-2.50	0.87	0.00	-2.50	0.87	0.00
C	-1.51	2.04	0.00	-1.51	2.04	0.00
C	-3.39	0.96	1.25	-3.39	0.96	1.25
C	-3.32	0.95	-1.16	-3.32	0.95	-1.16
O	-0.78	2.04	1.22	-2.22	3.27	-0.05
O	-4.14	2.16	1.23	-2.58	0.82	2.41
C	-4.30	-0.05	-1.41	-2.74	0.81	-2.45
C	-0.78	3.20	2.05	-1.51	4.50	-0.13
C	-5.10	2.43	2.24	-2.77	1.70	3.52
C	-3.96	-1.21	-1.54	-2.22	-0.24	-2.77
O	-0.34	4.25	1.63	-0.74	4.82	0.76
O	-4.75	2.50	3.41	-3.83	1.67	4.14
C	-5.76	0.33	-1.5	-2.77	1.97	-3.42
C	-1.33	3.11	3.46	-1.70	5.41	-1.33
C	-6.55	2.61	1.88	-1.68	2.66	3.93
C	-5.96	1.37	-2.61	-4.22	2.38	-3.70
C	-0.52	2.09	4.27	-3.18	5.80	-1.43
C	-6.69	3.79	0.89	-0.42	1.86	4.33
C	-7.45	1.71	-2.74	-4.25	3.51	-4.74
C	-1.16	1.91	5.65	-3.38	6.76	-2.62
C	-8.17	3.98	0.53	0.70	2.84	4.71
C	-7.64	2.76	-3.84	-5.71	3.91	-5.02
C	-0.37	0.86	6.45	-4.86	7.15	-2.73
C	-8.30	5.14	-0.47	1.94	2.05	5.15
C	-9.14	3.08	-4.00	-5.73	5.03	-6.07
C	-0.98	0.71	7.85	-5.05	8.14	-3.89
C	-9.78	5.34	-0.83	3.07	3.02	5.52
C	-9.33	4.14	-5.10	-7.19	5.42	-6.37
C	-0.21	-0.36	8.64	-6.53	8.52	-4.01
C	-9.91	6.51	-1.83	4.31	2.23	5.97
C	-10.82	4.45	-5.26	-7.22	6.53	-7.43
C	-0.80	-0.49	10.05	-6.71	9.55	-5.13

| | No. 1 (α form) | | | No. 3 (β form) | | |
	x	y	z	x	y	z
C	-11.40	6.71	-2.18	5.44	3.21	6.31
C	-11.01	5.51	-6.35	-8.68	6.92	-7.73
C	-0.04	-1.57	10.83	-8.20	9.92	-5.26
O	-11.53	7.88	-3.17	6.67	2.42	6.80
O	-12.51	5.81	-6.52	-8.71	8.02	-8.80
C	-0.62	-1.69	12.25	-8.37	10.97	-6.37
C	-13.01	8.10	-3.51	7.81	3.39	7.13
O	-12.70	6.86	-7.63	-10.16	8.39	-9.11
C	0.11	-2.80	13.01	-9.86	11.34	-6.49
C	-13.15	9.28	-4.48	9.03	2.60	7.63
O	-14.20	7.15	-7.81	-10.19	9.49	-10.19
O	-0.47	-2.91	14.43	-10.03	12.40	-7.59
C	-14.63	9.50	-4.82	10.18	3.57	7.94
C	-14.39	8.20	-8.91	-11.65	9.85	-10.51
C	0.25	-4.04	15.19	-11.51	12.79	-7.71
C	-14.77	10.68	-5.79	11.39	2.79	8.47
C	-15.89	8.47	-9.11	-11.68	10.97	-11.56
C	-0.34	-4.17	16.60	-11.68	13.86	-8.79
C	-16.25	10.91	-6.12	12.55	3.76	8.75
C	-16.08	9.51	-10.22	-13.14	11.32	-11.90
C	0.35	-5.32	17.35	-13.18	14.18	-8.97
C	-16.39	12.10	-7.07	13.74	2.98	9.32
C	-17.58	9.77	-10.43	-13.17	12.46	-12.93
C	-0.23	-5.43	18.77	-13.34	15.27	-10.03
C	-17.87	12.34	-7.39	14.91	3.94	9.58
C	-17.77	10.84	-11.51	-14.62	12.77	-13.31
C	0.46	-6.59	19.51	-14.84	15.54	-10.26
C	-18.01	13.54	-8.34	16.10	3.16	10.15
C	-19.27	11.11	-11.72	-14.66	13.93	-14.32
C	-0.13	-6.71	20.92	-15.01	16.66	-11.30
C	-19.49	13.75	-8.69	17.28	4.12	10.37

TABLE IV. Cartesian coordinates of the heavy atoms for the selected conformers of trielaidin.

	No. 14 (α form)			No. 4 (β form)			No. 14 (α form)			No. 4 (β form)		
	x	y	z	x	y	z	x	y	z	x	y	z
C	-2.50	0.87	0.00	-2.50	0.86	0.00	-1.19	-5.90	7.75	5.3	3.69	6.26
C	-1.51	2.04	0.00	-1.51	2.04	0.00	-11.06	5.15	-6.54	-8.72	6.72	-7.84
C	-3.39	0.96	1.25	-3.39	0.96	1.25	4.66	-6.42	-2.62	-8.18	9.83	-5.36
O	-3.32	0.95	-1.16	-3.32	0.95	-1.16	0.17	-6.10	8.38	6.57	2.99	6.71
O	-1.44	2.61	-1.30	-2.22	3.27	-0.06	-12.30	5.46	-6.94	-9.00	7.65	-8.76
O	-4.32	-0.11	1.25	-2.58	0.83	2.41	5.81	-7.11	-2.69	-8.57	10.69	-6.31
C	-4.23	-0.10	-1.46	-2.74	0.79	-2.45	0.37	-7.11	9.24	7.64	3.71	7.05
C	-0.98	1.84	-2.39	-1.51	4.50	-0.16	-12.51	6.56	-7.96	-10.44	8.02	-9.05
O	-5.24	-0.26	2.32	-2.83	1.65	3.55	5.79	-8.62	-2.74	-10.03	11.06	-6.43
O	-3.82	-1.24	-1.6	-2.22	-0.27	-2.75	1.73	-7.31	9.87	8.91	3.01	7.50
O	-1.72	1.59	-3.33	-0.74	4.83	0.73	-14.00	6.68	-8.29	-10.48	9.10	-10.14
C	-6.02	0.63	2.59	-3.87	1.53	4.17	7.22	-9.15	-2.86	-10.21	12.08	-7.57
C	-5.71	0.20	-1.68	-2.78	1.93	-3.44	1.67	-8.49	10.86	9.99	4.06	7.80
C	0.44	1.31	-2.39	-1.7	5.40	-1.36	-14.22	7.82	-9.29	-11.94	9.46	-10.45
C	-5.24	-1.54	3.14	-1.8	2.68	3.97	7.20	-10.69	-2.93	-11.69	12.46	-7.69
C	-5.92	1.22	-2.73	-4.24	2.31	-3.73	3.06	-8.71	11.47	11.27	3.35	8.28
C	0.42	-0.23	-2.44	-3.17	5.83	-1.44	-15.71	7.91	-9.66	-11.98	10.57	-11.51
C	-3.87	-1.73	3.8	-0.48	1.97	4.33	8.63	-11.22	-3.08	-11.87	13.49	-8.81
C	-7.43	1.48	-2.91	-4.28	3.41	-4.81	2.99	-9.87	12.48	12.37	4.39	8.53
C	1.86	-0.77	-2.46	-3.38	6.72	-2.6	-15.93	9.09	-10.62	-13.45	10.90	-11.86
C	-3.88	-3.01	4.65	0.57	3.02	4.73	8.61	-12.75	-3.12	-13.36	13.81	-8.99
C	-7.64	2.54	-4.00	-5.74	3.80	-5.09	4.39	-10.13	13.06	13.63	3.68	9.07
C	1.83	-2.30	-2.50	-4.86	7.12	-2.78	-17.41	9.15	-11.02	-13.49	12.03	-12.89
C	-2.51	-3.20	5.32	1.89	2.31	5.09	10.04	-13.28	-3.31	-13.53	14.87	-10.08
C	-9.14	2.77	-4.21	-5.79	4.88	-6.18	4.31	-11.26	14.10	14.74	4.72	9.31
C	3.27	-2.84	-2.51	-5.05	8.07	-3.97	-17.65	10.35	-11.95	-14.95	12.34	-13.26
C	-2.53	-4.46	6.19	2.95	3.35	5.46	10.01	-14.82	-3.35	-15.03	15.16	-10.29
C	-9.35	3.86	-5.28	-7.25	5.27	-6.46	5.72	-11.53	14.67	15.99	4.02	9.84
C	3.25	-4.38	-2.53	-6.53	8.48	-4.07	-19.12	10.39	-12.39	-14.99	13.50	-14.26
C	-1.16	-4.65	6.86	4.24	2.64	5.88	11.44	-15.35	-3.58	-15.20	16.28	-11.33
C	-10.85	4.05	-5.52	-7.29	6.35	-7.55	5.63	-12.63	15.74	17.10	5.06	10.07
C	4.69	-4.91	-2.57	-6.72	9.46	-5.24						
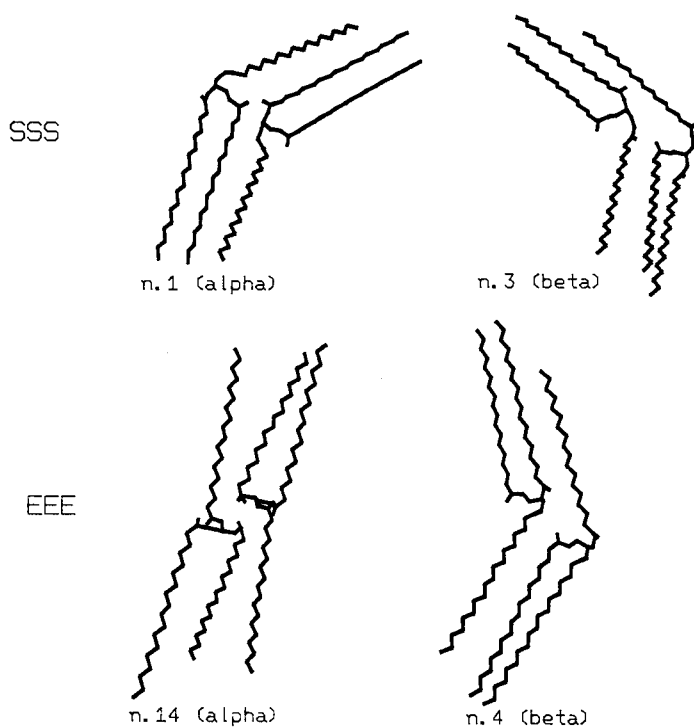
Figure 11. Bimolecular assembly of the select conformers of tristearin (SSS) and trielaidin (EEE).

were generally obtained) for which the structure was solved at low temperature [7]. Moreover, this method can be applied only to stable polymorphic forms.

For these major reasons, to compensate for the lack of information coming from monocrystals X-ray diffraction, we developed a molecular modeling technique,

TABLE V. Value of interaction energy (modelization results) and of ΔH (differential scanning calorimetry measurements) determined for conformers of trilaurin (LLL), tristearin (SSS), and trielaidin (EEE).

Experimental ΔH (kcal/mol)	Conformer	E_{int} (kcal/mol)
	LLL no. 1	−35.3
	SSS	
25.5	(α) no. 1	−157.5
44.0	(β) no. 3	−208.5
	EEE	
9.7	(α) no. 14	−67.9
32.9	(β) no. 4	−110.9

CULOT ET AL.

Figure 12.   Progressive assembly in monolayer for trielaidin under the (top) α form and (bottom) β form.

allowing us to simulate the three-dimensional structures of the different, stable or less stable, polymorphic forms of triglycerides. First, isolated molecule conformers can be generated by a systematic structure-tree analysis followed by a minimization procedure. The total conformational energy includes van der Waals, Coulomb, and torsional contributions. Then, selected monomers can be assembled into head-to-tail dimers and later in monolayers. Therefore, the interaction energy is computed as a sum of van der Waals, Coulomb, and transfer energy terms. Because of the importance of the electrostatic contribution in the evaluation of both the conformational energy of the isolated molecules as well as for the interaction energy of the assembly of the dimers, the atomic charge calculations were handled at the *ab initio* MO–SCF level with the 6-31G basis set.

From a detailed study of trilaurin, a C_{12} triglyceride, one of the only triglycerides whose structure has been solved, a work strategy has been defined and applied to the conformational analysis of C_{18} monoacid triglycerides, i.e., tristearin, trielaidin, and triolein.

For the generation of the isolated conformers, with the proposed structure-tree analysis (variations around 11 torsional angles along the three chains) and the proposed potential functions (van der Waals, torsional, and electrostatic terms), one can use atomic charges obtained by the Mulliken population analysis. For the assembly of dimers and monolayers, one needs to use the potential-derived charges previously calculated for each conformer.

The structure-tree analysis generated many low-energy conformers: 12 for tristearin, 14 for trielaidin, and 12 for triolein. For triolein, interestingly, one obtains preferentially folded structures, as found in biological systems. Tristearin and trielaidin presented mostly extended structures. For these, and by correlation with experimental data (NMR, powder X-ray diffraction), conformers representing α and β forms were retained.

The selected conformers were then sequentially assembled in head-to-tail bimolecular configurations. For each assembly, interaction energy values were calculated and compared to enthalpy values, measured experimentally. Results obtained by modelization, concerning the stability of the different assemblies, correlate well with experimental observations: The β form is more stable than is the α form, and the presence of *trans* unsaturations lowers the compacity of the lattice.

For monoacid triglycerides, molecular modeling can thus reproduce the structures and packing of the polymorphic forms as well as the effects of chain length and the presence of unsaturations on the stability of crystalline edifices. Moreover, it allows the simulation of less stable forms.

Presently, we are applying the same strategy to diacid triglycerides, mixed saturated and unsaturated, in order to underline, at a molecular level, the influence of *cis* and *trans* unsaturations on the polymorphism. This work is without doubt of technological interest, as such compounds are mostly present in natural fats.

Bibliography

[1] H. Baumann, M. Bülher, H. Fochem, F. Hirsinger, H. Zoebelein, and J. Falbe, Angew. Chem. **27**, 41 (1988).

[2] G. Assman, in *Lipid Metabolism and Atherosclerosis*, F. K. Schattauer, Ed. (Verlag GmbH, Stuttgart, Germany, 1982), p. 1.

[3] V. Bézard, Inf. Diétét. **34**, 25 (1983).

[4] S. Eisenberg, J. Lipid Res. **25**, 1017 (1984).

[5] E. J. Mc Murchie, J. A. Rinaldi, S. L. Burnard, G. S. Patten, M. Neumann, G. H. Mc Intosh, M. Abbey, and R. A. Gibson, Biochim. Biophys. Acta **1045**, 164 (1990).

[6] A. Desmedt, C. Culot, C. Deroanne, F. Durant, and V. Gibon, J. Am. Oil Chem. Soc. **67**, 635 (1990).

[7] C. Culot, PhD Thesis (University of Namur, Belgium, January 1994).

[8] D. M. Hegsted, in *Health Effects of Dietary Fatty Acids*, G. J. Nelson, Ed. (Am. Oil Chem. Soc., Champaign, IL, 1991), p. 50.

[9] K. Larsson, Ark. Kemi **23**, 1 (1964).

[10] L. H. Jensen and A. J. Mabis, Acta Crystallogr. **21**, 770 (1966).

[11] L. T. H. Doyne and J. T. Gordon, J. Am. Oil Chem. Soc. **45**, 333 (1968).

[12] M. Goto, D. R. Kodali, D. M. Small, K. Honda, K. Kozawa, and T. Uchida, Proc. Natl. Acad. Sci. U.S.A. **89**, 8083 (1992).

[13] R. Brasseur, in *Molecular Description of Biological Membranes by Computer Aided Conformational Analysis*, R. Brasseur, Ed. (CRC Press, Boca Raton, FL, 1990), p. 203.

[14] A. M. Liquori, E. Giglio, and L. Mazzarella, Nuovo Cimento **55B**, 475 (1968).

[15] E. Giglio, A. M. Liquori, and L. Mazzarella, Nuovo Cimento **56B**, 57 (1968).

[16] A. J. Hopfinger, in *Conformational Properties of Macromolecules*, B. Horecker, Ed. (Academic Press, New York, 1973), p. 118.

[17] R. Brasseur and M. Deleers, Proc. Natl. Acad. Sci. U.S.A. **81**, 3370 (1984).

[18] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).

[19] A. C. Tanford, *Hydrophobic Effect Formation of Micelles and Biological Membranes* (Wiley, New York, 1973), p. 1.

[20] R. Brasseur, J. Biol. Chem. **266**, 16120 (1991).

[21] C. Culot, M. Dory, F. Durant, and D. P. Vercauteren, Int. J. Quantum Chem. **46**, 211 (1993).

[22] J. A. Pople, in *Applications of Electronic Structure Theory*, Modern Theoretical Chemistry Series, Vol 4, H. F. Schaefer III, Ed. (Plenum, New York, 1977), p. 1.

[23] M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, GAUSSIAN 92, Revision 0.4 (Gaussian Inc., Pittsburgh, PA, 1992).

[24] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).

[25] P. A. Kollman, J. Am. Chem. Soc. **100**, 2974 (1978).
[26] U. C. Singh and P. A. Kollman, J. Comp. Chem. **5**, 129 (1984).
[27] C. Culot and F. Durant, in *Proceedings of the First International Conference on Applications of Magnetic Resonance in Food Science,* Guildford, England (1992), p. 3.
[28] D. M. Small, in *Handbook of Lipid Research, The Physical Chemistry of Lipids: From Alkanes to Phospholipids,* D. M. Small, Ed. (Plenum, New York, 1986), Vol. 4, p. 345.

Ab Initio SCF Investigation of the Intramolecular Hydrogen Bonding in ε-Aminohexanoic Acid

MICHAEL RAMEK

Institut für Physikalische und Theoretische Chemie, Technische Universität Graz,
A-8010 Graz, Austria

Abstract

Ab initio RHF results (geometry data and potential barriers) are reported for the 14 symmetry unique local minima in the potential energy surface of ε-aminohexanoic acid, which contain an intramolecular N · · · H — O hydrogen bond. Comparison of characteristic data with those of the homologs with fewer carbon atoms shows that δ-aminopentanoic acid forms the most stable hydrogen bonds. In contrast to these homologs, the — COOH group is not limited to *anti-periplanar* orientation in the H-bonded conformers of ε-aminohexanoic acid: In four such conformers, it occurs in *syn-clinal* orientation. © 1994 John Wiley & Sons, Inc.

Introduction

The intramolecular N · · · H — O hydrogen bonding in the neutral form of ω-amino acids has been the target of *ab initio* studies for a number of years. For glycine, one mirror symmetrical conformer with this H bond is found with split valence basis sets [1–3]; with basis sets that include polarization functions, the nature of this conformation changes to a transition state between two mutually mirror symmetrical local minima with slightly lower energy [4,5]. Polarized basis sets, however, apparently tend to produce unrealistic structures in such systems with intramolecular hydrogen bonds [6]. In β-alanine, one symmetry unique conformer with a N · · · H — O hydrogen bond is formed, which is connected with its mirror image by a reaction path that preserves the H bond [7]. In γ-aminobutyric acid (GABA), two mirror symmetrical conformers with this H bond are formed, each of which is connected with both, the image and the mirror image of its counterpart, in an H bond preserving reaction [8]. δ-Aminopentanoic acid follows this pattern: There are four symmetry unique H-bonded conformers that are interconnected in such a way that for each conformer three different H bond preserving reaction paths exist [9]. In all of these systems, the hydrogen bond can only be formed with the — COOH group in *anti-periplanar* orientation (i.e., the tosion angle H — O — C = O is close to 180°); this orientation is approximately 35 kJ/mol less stable than is the *syn-periplanar* orientation with a H — O — C = O torsion angle close to 0°. The stabilization due to the hydrogen bond is slightly lower than this energy difference, so none of the H-bonded conformers is the global minimum in the respective potential energy surface:

$$-C \overset{\displaystyle O}{\underset{\displaystyle \overset{O}{\underset{H}{\diagup}}}{\diagdown}} \qquad -C \overset{\displaystyle O}{\underset{\displaystyle O-H}{\diagdown}}$$
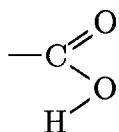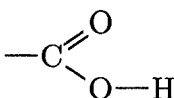
anti-periplanar   *syn-periplanar*

The strength of the $N \cdots H - O$ interaction increases monotonically from glycine to δ-aminopentanoic acid [3,9] with changes of several hundred cm^{-1} in the calculated $O - H$ vibration frequency. This dependence of the H-bond strength on the ring size and the interesting topological pattern of the H bond preserving reaction paths lead to an extension of this series of *ab initio* calculations to the next homolog, ε-aminohexanoic acid. The well-known RHF formalism [10] and the 4-31G [11] basis set were used in order to allow a comparison of the results with those of earlier work at one consistent level. All geometries were fully optimized with the program GAMESS [12] to remaining maximum and root mean square (rms) gradients less than 1×10^{-4} and 0.33×10^{-4} H Bohr^{-1}, respectively. Local minima were verified to have only positive eigenvalues of the Hessian matrix, and saddle points were verified to have exactly one negative eigenvalue of the Hessian matrix. The Hessian matrix was obtained via numerical differentiation of analytical first derivatives; no scaling was performed in the subsequent vibration frequency analysis.

Results

The potential energy surface of ε-aminohexanoic acid contains a total of 14 symmetry unique local minima that are stabilized by an intramolecular $N \cdots H - O$ hydrogen bond. In accordance with earlier work, the torsion angle $N - C - C - C$ was chosen as the criterion for symmetry uniqueness. Conformers with a positive value of this angle are labeled **I, II, . . ., XIV** in the following; conformers with a negative value are labeled **I^m, II^m, . . ., XIV^m**. The geometry data of these local minima are collected in Table I; characteristic data related to the H bond are summarized in Table II.

As a consequence of the intramolecular hydrogen bond, various internal rotations are coupled in most reaction paths of the H-bonded conformers. Table III lists the potential barriers for all of these reaction paths; reactions that preserve the H bond are also indicated in this table. In contrast to the homologs up to δ-aminopentanoic acid, the H bond preserving reactions of ε-aminohexanoic acid yield the very complex pattern that is displayed in Figure 1.

The two conformers **VIII** and **XIII** are noteworthy because of the small value of the lowest potential barrier. In **VIII**, this value is 0.90 kJ/mol and corresponds to a combined rotation of the groups $-OH$, $-COOH$, and $-NH_2$ that destroys the $N \cdots H - O$ interaction. The harmonic vibration frequency of this mode in **VIII** is 86.9 cm^{-1}, which is equivalent to a vibrational zero-point energy of 0.52 kJ/mol. Contrastingly, in **XIII**, the lowest potential barrier of 0.06 kJ/mol is caused by a H bond preserving reaction; the harmonic frequency of **XIII** associated with this mode is 40.2 cm^{-1}, which gives a zero-point energy of 0.24 kJ/mol. Even if

TABLE I.   Geometry data and energies of all symmetry unique conformers in the potential energy surface of ε-aminohexanoic acid with a N · · · H—O hydrogen bond.

	Conformer				
	I	II	III	IV	V
SCF energy (au)	−438.3175141	−438.3173938	−438.3160495	−438.3158663	−438.3156863
Bond lengths (Å)					
C6—N	1.4737	1.4716	1.4693	1.4716	1.4765
C5—C6	1.5277	1.5291	1.5327	1.5292	1.5342
C4—C5	1.5455	1.5364	1.5334	1.5369	1.5381
C3—C4	1.5353	1.5393	1.5361	1.5406	1.5319
C2—C3	1.5432	1.5459	1.5266	1.5263	1.5439
C1—C2	1.5100	1.5096	1.5194	1.5179	1.5071
H1—N	1.0008	1.0004	1.0005	1.0001	1.0014
H2—N	1.0003	1.0011	0.9998	1.0004	1.0013
H4—C6	1.0861	1.0867	1.0864	1.0867	1.0860
H5—C6	1.0815	1.0796	1.0804	1.0818	1.0817
H6—C5	1.0859	1.0871	1.0859	1.0869	1.0850
H7—C5	1.0842	1.0858	1.0855	1.0846	1.0847
H8—C4	1.0840	1.0816	1.0824	1.0842	1.0860
H9—C4	1.0823	1.0851	1.0850	1.0847	1.0859
H10—C3	1.0838	1.0854	1.0820	1.0843	1.0812
H11—C3	1.0849	1.0838	1.0842	1.0811	1.0834
H12—C2	1.0795	1.0795	1.0840	1.0822	1.0783
H13—C2	1.0822	1.0798	1.0831	1.0851	1.0836
O1—C1	1.2065	1.2063	1.2068	1.2068	1.2044
O2—C1	1.3372	1.3376	1.3359	1.3356	1.3418
H3—O2	0.9748	0.9753	0.9719	0.9727	0.9718
Valence angles (degree)					
C5—C6—N	111.07	111.92	111.54	112.30	112.12
C4—C5—C6	113.73	117.41	114.15	113.39	115.20
C3—C4—C5	114.44	116.66	114.51	113.83	115.71
C2—C3—C4	114.89	117.21	112.98	113.54	113.85
C1—C2—C3	112.13	112.46	116.89	116.53	112.24
H1—N—C6	112.85	113.36	113.82	113.43	112.45
H2—N—C6	113.52	113.19	113.68	113.25	112.78
H4—C6—C5	109.87	108.30	109.01	109.45	109.35
H5—C6—C5	109.78	110.59	110.17	109.03	110.20
H6—C5—C4	110.56	107.86	108.40	109.96	110.60
H7—C5—C4	109.63	109.14	110.32	109.90	107.52
H8—C4—C3	108.03	110.13	109.66	108.38	108.22
H9—C4—C3	109.26	107.22	108.94	110.13	109.51
H10—C3—C4	109.41	109.44	110.80	108.91	109.00
H11—C3—C4	109.31	108.23	109.26	110.41	109.01
H12—C2—C3	108.97	111.34	110.24	111.03	110.87
H13—C2—C3	110.57	108.34	110.87	110.03	110.23
O1—C1—C2	122.33	122.64	120.21	120.61	124.10
O2—C1—C2	116.65	116.14	119.62	119.04	114.62
H3—O2—C1	118.40	117.38	123.16	122.63	116.64

TABLE I.   *(Continued)*

	Conformer				
	I	**II**	**III**	**IV**	**V**
Torsion angles (degree)					
C4—C5—C6—N	71.29	53.20	43.16	58.33	90.77
C3—C4—C5—C6	−128.26	63.28	53.70	−90.15	−54.79
C2—C3—C4—C5	69.22	−82.02	−160.38	154.24	−66.47
C1—C2—C3—C4	67.55	−59.31	70.15	−66.31	149.63
H1—N—C6—C5	72.43	178.94	60.21	66.40	64.76
H2—N—C6—C5	−161.76	52.94	−172.82	−167.27	−170.83
H4—C6—C5—C4	−165.41	176.30	167.35	−177.23	−146.36
H5—C6—C5—C4	−47.72	−67.09	−75.81	−60.22	−29.12
H6—C5—C4—C3	−5.91	−174.21	175.70	33.28	69.19
H7—C5—C4—C3	110.33	−59.45	−68.29	149.77	−175.81
H8—C4—C3—C2	−169.22	44.40	−37.87	−83.42	171.95
H9—C4—C3—C2	−54.52	158.81	77.48	32.00	56.56
H10—C3—C4—C5	−169.05	41.44	−37.95	−84.71	56.82
H11—C3—C4—C5	−52.55	157.06	78.42	31.03	172.20
H12—C2—C3—C4	−174.28	64.16	−169.13	55.77	−89.18
H13—C2—C3—C4	−55.73	−177.36	−52.15	172.89	29.99
O1—C1—C2—C3	92.59	−92.18	−176.19	167.80	118.47
O2—C1—C2—C3	−85.93	86.64	5.47	−14.29	−61.86
H3—O2—C1—C2	1.58	−3.29	−9.86	9.12	−0.42
Distances (Å) less than 95.0% of sum of van der Waals radii					
H3—N	1.8213	1.8207	1.8726	1.8517	1.8665
H5—H12		2.1170			
H5—H10			2.1837		
H3—H8			2.1997		
H3—H10			2.1895		2.2080
H3—H11				2.1224	
H13—H6					2.2329
Additional attractive interactions (Å)					
O2—H2				2.9640	
H8—N		2.7782	2.6570		
H9—N	2.8362			2.9937	
H10—N			2.7793		2.7724
H11—N				2.5664	
H12—N		2.9250			
O2—H8		2.5386	2.6397		
O2—H9	2.5790			2.7043	
O2—H10			2.7021		2.5618
O2—H11				2.7024	
Vibrational zero-point energies (kJ/mol)					
	553.710	554.178	552.516	552.316	553.624
Rotation constants (GHz)					
	0.8339	0.8777	0.7818	0.8018	0.7792
	1.0797	1.0968	0.9628	0.9868	1.0086
	2.4041	2.4186	2.7581	2.6078	2.5160

TABLE I. *(Continued)*

	Conformer				
	VI	**VII**	**VIII**	**IX**	**X**
SCF energy (au)	−438.3148969	−438.3146779	−438.3134942	−438.3131759	−438.3124873
Bond lengths (Å)					
C6—N	1.4754	1.4778	1.4768	1.4757	1.4773
C5—C6	1.5404	1.5370	1.5330	1.5331	1.5287
C4—C5	1.5416	1.5508	1.5400	1.5372	1.5470
C3—C4	1.5410	1.5459	1.5514	1.5368	1.5370
C2—C3	1.5396	1.5545	1.5523	1.5462	1.5485
C1—C2	1.4978	1.4938	1.4924	1.5052	1.5124
H1—N	0.9991	1.0016	1.0003	1.0010	0.9993
H2—N	0.9992	0.9995	1.0001	1.0012	0.9988
H4—C6	1.0865	1.0863	1.0874	1.0860	1.0811
H5—C6	1.0785	1.0817	1.0790	1.0808	1.0861
H6—C5	1.0851	1.0824	1.0875	1.0867	1.0803
H7—C5	1.0801	1.0845	1.0862	1.0848	1.0843
H8—C4	1.0867	1.0842	1.0791	1.0843	1.0841
H9—C4	1.0858	1.0806	1.0845	1.0858	1.0863
H10—C3	1.0848	1.0831	1.0842	1.0825	1.0839
H11—C3	1.0804	1.0839	1.0830	1.0825	1.0842
H12—C2	1.0849	1.0793	1.0799	1.0812	1.0824
H13—C2	1.0804	1.0798	1.0795	1.0781	1.0800
O1—C1	1.2096	1.2129	1.2110	1.2038	1.2072
O2—C1	1.3433	1.3404	1.3451	1.3416	1.3354
H3—O2	0.9734	0.9801	0.9758	0.9709	0.9770
Valence angles (degree)					
C5—C6—N	110.93	111.94	113.27	112.88	112.08
C4—C5—C6	115.28	115.35	118.16	117.58	114.27
C3—C4—C5	117.43	113.70	115.16	115.75	115.65
C2—C3—C4	116.13	114.71	115.07	113.84	115.85
C1—C2—C3	112.73	109.66	108.49	111.44	114.27
H1—N—C6	113.83	112.73	112.47	112.42	113.47
H2—N—C6	114.04	113.24	113.36	112.43	113.85
H4—C6—C5	109.53	109.30	107.43	108.45	109.04
H5—C6—C5	110.64	110.33	111.02	110.28	110.02
H6—C5—C4	107.61	110.47	108.26	109.10	110.63
H7—C5—C4	109.76	108.24	108.58	107.46	109.36
H8—C4—C3	107.66	108.67	110.10	108.62	107.66
H9—C4—C3	108.61	109.24	108.47	109.76	108.31
H10—C3—C4	107.40	109.72	110.39	109.67	107.78
H11—C3—C4	110.47	109.44	109.51	109.01	110.20
H12—C2—C3	108.47	109.67	109.38	110.17	110.18
H13—C2—C3	111.64	109.60	110.14	110.48	108.24
O1—C1—C2	124.76	124.76	124.29	124.44	121.53
O2—C1—C2	112.38	112.19	112.68	114.01	117.50
H3—O2—C1	115.17	113.98	113.15	115.18	118.44

TABLE I. *(Continued)*

	Conformer				
	VI	**VII**	**VIII**	**IX**	**X**
Torsion angles (degree)					
C4—C5—C6—N	138.89	90.10	66.32	83.79	60.14
C3—C4—C5—C6	−62.61	−144.59	64.30	−65.82	−124.41
C2—C3—C4—C5	−61.97	90.86	−117.35	92.47	71.90
C1—C2—C3—C4	64.96	−43.79	59.79	−150.17	−87.59
H1—N—C6—C5	175.79	62.47	174.08	57.09	149.83
H2—N—C6—C5	47.43	−171.25	47.95	−178.86	−83.36
H4—C6—C5—C4	−97.49	−147.28	−171.34	−153.77	178.92
H5—C6—C5—C4	20.47	−29.81	−54.75	−37.12	−63.75
H6—C5—C4—C3	175.56	−21.44	−172.23	58.66	−1.76
H7—C5—C4—C3	61.26	93.84	−57.68	173.10	114.76
H8—C4—C3—C2	174.99	−148.56	7.36	−147.40	−166.13
H9—C4—C3—C2	60.50	−32.76	123.21	−31.93	−52.62
H10—C3—C4—C5	177.73	−148.48	5.23	−146.62	−168.20
H11—C3—C4—C5	61.93	−31.79	121.83	−30.79	−52.59
H12—C2—C3—C4	−175.49	75.59	179.15	−30.23	37.57
H13—C2—C3—C4	−56.78	−164.32	−60.43	89.01	155.16
O1—C1—C2—C3	−129.69	−85.51	86.42	−109.61	−85.95
O2—C1—C2—C3	48.53	88.96	−87.48	69.38	93.73
H3—O2—C1—C2	−140.60	−138.63	131.03	2.30	−9.49
Distances (Å) less than 95.0% of sum of van der Waals radii					
H3—N	1.9468	1.8283	1.8900	1.8679	1.7942
H5—H11	2.2309				
O2—H9		2.4263			
O2—H8			2.3567		
Additional attractive interactions (Å)					
H8—N			2.9390		
H9—N					2.7184
H11—N				2.9718	
H12—N				2.8558	
O1—H5			2.6010		
O2—H5	2.6737				
O1—H6		2.5998			
O2—H6					2.5297
O1—H7	2.5686				
O2—H7	2.8328				
O1—H11					2.9874
O2—H11	2.5210			2.6356	
Vibrational zero-point energies (kJ/mol)					
	551.663	552.594	551.547	553.607	553.201
Rotation constants (GHz)					
	1.0246	1.0057	1.0392	0.7805	0.9052
	1.4497	1.3803	1.4313	1.0053	1.1678
	1.8630	2.0304	1.9354	2.5159	2.1733

TABLE I. *(Continued)*

	Conformer			
	XI	**XII**	**XIII**	**XIV**
SCF energy (au)	−438.3124520	−438.3123593	−438.3050339	−438.3032209
Bond lengths (Å)				
C6—N	1.4781	1.4760	1.4793	1.4701
C5—C6	1.5466	1.5342	1.5407	1.5397
C4—C5	1.5540	1.5388	1.5481	1.5473
C3—C4	1.5467	1.5396	1.5430	1.5397
C2—C3	1.5409	1.5335	1.5438	1.5438
C1—C2	1.4932	1.5113	1.5081	1.5132
H1—N	0.9998	1.0009	1.0014	0.9998
H2—N	0.9997	0.9983	0.9995	0.9981
H4—C6	1.0862	1.0849	1.0852	1.0856
H5—C6	1.0772	1.0825	1.0795	1.0806
H6—C5	1.0813	1.0827	1.0804	1.0851
H7—C5	1.0839	1.0852	1.0837	1.0842
H8—C4	1.0841	1.0857	1.0839	1.0856
H9—C4	1.0862	1.0864	1.0852	1.0852
H10—C3	1.0845	1.0820	1.0844	1.0832
H11—C3	1.0818	1.0846	1.0828	1.0786
H12—C2	1.0801	1.0871	1.0781	1.0837
H13—C2	1.0842	1.0803	1.0816	1.0782
O1—C1	1.2123	1.2065	1.2055	1.2051
O2—C1	1.3392	1.3369	1.3406	1.3406
H3—O2	0.9792	0.9726	0.9716	0.9721
Valence angles (degree)				
C5—C6—N	111.95	112.54	110.47	113.71
C4—C5—C6	116.18	114.21	117.62	119.63
C3—C4—C5	116.12	117.67	119.78	119.66
C2—C3—C4	116.59	115.75	118.48	116.55
C1—C2—C3	112.32	118.84	113.85	113.87
H1—N—C6	113.63	113.17	112.68	113.64
H2—N—C6	113.02	114.09	113.80	114.73
H4—C6—C5	109.40	109.56	110.19	107.85
H5—C6—C5	110.97	109.60	111.21	110.22
H6—C5—C4	108.88	111.46	108.76	106.34
H7—C5—C4	108.02	107.18	107.03	109.24
H8—C4—C3	106.76	108.74	108.34	106.95
H9—C4—C3	109.86	107.87	106.25	109.03
H10—C3—C4	108.24	110.56	109.73	107.55
H11—C3—C4	109.10	107.54	106.68	108.53
H12—C2—C3	111.08	108.48	110.46	110.92
H13—C2—C3	108.77	110.33	108.08	111.44
O1—C1—C2	124.29	121.33	123.03	122.57
O2—C1—C2	112.96	118.09	115.61	116.76
H3—O2—C1	113.36	121.10	115.81	119.57

Table I.   *(Continued)*

	Conformer			
	XI	**XII**	**XIII**	**XIV**
Torsion angles (degree)				
C4—C5—C6—N	128.71	87.57	113.72	31.19
C3—C4—C5—C6	−102.61	−60.94	−33.79	39.59
C2—C3—C4—C5	86.46	−83.44	−43.46	41.16
C1—C2—C3—C4	−52.94	67.85	−42.14	−131.07
H1—N—C6—C5	61.10	86.18	69.18	165.13
H2—N—C6—C5	−171.71	−146.78	−164.81	36.55
H4—C6—C5—C4	−107.92	−149.36	−123.52	155.29
H5—C6—C5—C4	9.71	−32.56	−5.50	−88.99
H6—C5—C4—C3	20.54	63.58	89.14	162.84
H7—C5—C4—C3	135.67	179.21	−157.05	−83.81
H8—C4—C3—C2	−153.58	40.67	79.49	165.35
H9—C4—C3—C2	−39.37	155.20	−167.12	−81.20
H10—C3—C4—C5	−153.24	43.32	79.78	163.25
H11—C3—C4—C5	−37.41	158.20	−165.62	−83.13
H12—C2—C3—C4	67.72	−171.69	81.58	−10.30
H13—C2—C3—C4	−173.43	−55.48	−161.10	108.19
O1—C1—C2—C3	124.60	−151.07	−72.40	−136.38
O2—C1—C2—C3	−52.41	33.20	107.32	45.63
H3—O2—C1—C2	143.61	−21.16	−9.26	−3.24
Distances (Å) less than 95.0% of sum of van der Waals radii				
H3—N	1.8612	1.8229	1.8972	1.8551
H3—H10		2.1595		
H3—H6			2.2709	
H3—H12			2.1457	
H3—H11				2.0412
Additional attractive interactions (Å)				
O2—H1		2.7576		
H10—N		2.6095		
H11—N				2.6627
H12—N			2.6580	
O1—H5	2.5088			
O2—H5				2.7890
O2—H6	2.6097	2.6505	2.5409	
O2—H10		2.7797		
O2—H11	2.6741			2.4741
Vibrational zero-point energies (kJ/mol)				
	552.122	551.661	553.961	553.571
Rotation constants (GHz)				
	1.0357	0.9093	0.9843	0.8693
	1.5181	1.1179	1.2592	1.0792
	1.8666	2.2306	2.0800	2.3979

TABLE II. Characteristic data of the N···H—O hydrogen-bonded conformers: SCF energy (au), N···H distance (Å) and bond order [14], O—H distance (Å) and harmonic frequency (cm^{-1}), H—O—C=O torsion angle (degree), N···H—O angle (degree), and electron density in critical point along the N···H connection line.

| Conformer | Energy | N···H | | O—H | | | | |
		Distance	Bond order	Distance	Frequency	H—O—C=O	N···H—O	Density
I	-438.3175141	1.8213	0.094	0.9748	3501	-177.13	175.36	0.04302
II	-438.3173938	1.8207	0.093	0.9753	3495	167.55	176.56	0.04324
III	-438.3160495	1.8726	0.087	0.9719	3570	171.80	162.19	0.03841
IV	-438.3158663	1.8517	0.083	0.9727	3549	-172.96	163.07	0.04007
V	-438.3156863	1.8665	0.093	0.9718	3574	179.27	174.71	0.03874
VI	-438.3148969	1.9468	0.084	0.9734	3607	37.66	139.89	0.03423
VII	-438.3146779	1.8283	0.101	0.9801	3463	35.97	152.09	0.04317
VIII	-438.3134942	1.8900	0.091	0.9758	3561	-42.99	147.22	0.03831
IX	-438.3131759	1.8679	0.092	0.9709	3594	-178.69	177.14	0.03859
X	-438.3124873	1.7942	0.094	0.9770	3454	170.18	171.15	0.04591
XI	-438.3124520	1.8612	0.097	0.9792	3489	-33.45	146.20	0.04076
XII	-438.3123593	1.8229	0.091	0.9726	3545	163.07	155.40	0.04239
XIII	-438.3050339	1.8972	0.089	0.9716	3583	170.47	164.92	0.03708
XIV	-438.3032209	1.8551	0.079	0.9721	3561	178.73	167.00	0.03945

TABLE III.   Potential barriers (kJ/mol) of all reaction paths of conformers
I–XIV.[a]

Conformer	Potential barrier	Reaction path description
I	22.93	2;
	26.47	10, 11, 14; leads to **II**[m]
	28.11	4, 5, 8, 10; leads to **IV**
	36.71	1, 4, 9;
	42.89	6, 7, 10; leads to **V**
	44.41	3, 12;
	53.20	9;
II	30.04	3, 6, 7; leads to **III**
	32.51	8, 9, 12; leads to **XIII**
	32.61	1, 4;
	33.24	10, 11, 14; leads to **I**[m]
	36.38	3, 6, 9, 12; leads to **XII**
	37.19	2, 3;
	43.58	5, 8; leads to **XIV**
III	11.82	1, 4;
	17.80	2, 3; leads to **VIII**
	21.39	10, 11, 13; leads to **IV**[m]
	26.52	4, 5, 8; leads to **II**
	26.56	8, 9, 12; leads to **XII**
	28.39	3, 6, 8, 9, 12; leads to **V**
	34.60	3, 8, 10, 11; leads to **X**[m]
	42.17	7, 4;
IV	15.53	4, 5, 7, 12; leads to **IX**
	19.82	2, 7, 12; leads to **XI**
	20.66	4, 7, 9; leads to **X**
	20.91	10, 11, 13; leads to **III**[m]
	23.79	3, 6, 7; leads to **I**
	24.69	1, 4, 7, 9, 12; leads to **VII**
	45.24	8;
V	10.02	1, 4, 5; leads to **XII**
	27.43	7, 10, 11; leads to **III**
	29.94	2;
	34.50	3;
	38.09	5, 8, 9, 11; leads to **I**
	40.15	6, 12;
	44.54	6;
VI	2.45	1, 7, 12;
	13.97	2, 7, 11; leads to **XII**
	23.42	1, 4, 5, 12;
	33.18	1, 8, 10, 11;
	34.83	8, 9, 11;
VII	5.92	1, 4, 8, 9, 11;
	6.70	1, 9, 11;
	9.15	1, 3, 6, 7, 9, 11;
	12.31	1, 4, 5, 10, 12;
	17.88	2, 5, 10, 11; leads to **X**
	21.57	2, 3, 8, 10, 11; leads to **IV**

TABLE III. *(Continued)*

Conformer	Potential barrier	Reaction path description
VIII	0.90	2, 3, 7;
	11.09	1, 4; leads to **III**
	17.70	2, 5, 8, 4;
	24.24	3, 8, 9, 12;
	35.23	2, 7, 10, 11;
IX	9.00	3, 6, 8, 11; leads to **IV**
	14.83	6, 9, 11; leads to **X**
	23.73	6, 7, 10, 12; leads to **XIII**
	29.19	6, 7, 10, 11; leads to **XIV**
	30.66	2, 3, 6, 9, 12; leads to **XI**
	44.25	4, 7, 9, 12;
	44.60	5, 9;
X	11.79	2, 3, 8, 10; leads to **IV**
	12.13	1, 6, 9, 12; leads to **VII**
	13.02	5, 10; leads to **IX**
	20.96	4;
	24.97	9;
	25.25	10, 11; leads to **III**^{m}
XI	5.85	2, 5, 7, 9, 12;
	10.62	1, 8, 11; leads to **IV**
	25.88	2, 3, 6, 7, 10, 12;
	28.76	1, 4, 5, 10, 11; leads to **IX**
XII	1.29	2, 3, 6; leads to **V**
	7.30	1, 8, 12; leads to **VI**
	16.87	7, 10, 11; leads to **III**
	23.16	4, 5, 10, 11; leads to **II**
	23.16	4, 5, 8, 10, 12; leads to **XIII**
	26.53	2, 3, 6, 11; leads to **XIV**^{m}
XIII	0.06	7, 10, 11; leads to **II**
	2.36	5, 8, 9, 11; leads to **IX**
	3.93	3, 6, 7, 9, 11; leads to **XII**
XIV	2.53	2, 4, 6, 10, 11; leads to **XII**^{m}
	3.06	5, 8, 9, 12; leads to **IX**
	6.37	6, 7; leads to **II**
	24.46	1, 5;

[a] In most of these reactions, several internal rotations are coupled as a consequence of the intramolecular N · · · H—O hydrogen bond. Internal rotations are labeled as follows: 1, 2, change of dihedral H—O—C1—C2; 3, 4, change of dihedral O=C1—C2—C3; 5, 6, change of dihedral C1—C2—C3—C4; 7, 8, change of dihedral C2—C3—C4—C5; 9, 10, change of dihedral C1—C2—C3—C4; 11, 12, change of dihedral C4—C5—C6—N; 13, 14, change of dihedrals C5—C6—N—H; odd/even numbers indicate decrease/increase of the respective dihedral angle. For reactions that preserve the hydrogen bond, the target conformer is also listed.
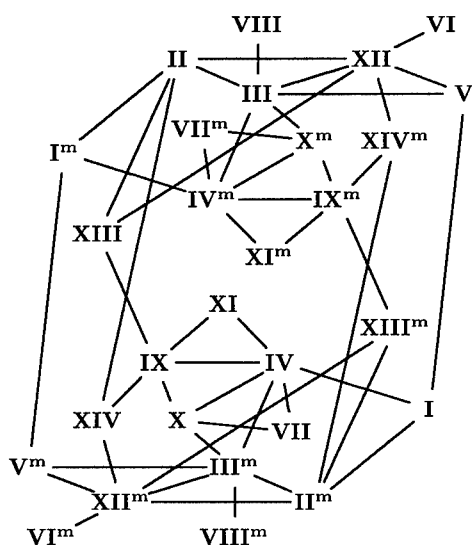
Figure 1. All reaction paths between conformers I–XIV and their mirror images I^m–XIV^m that preserve the $N \cdots H — O$ hydrogen bond.

an extraordinary degree of anharmonicity is assumed, the zero-point energy still is larger than the potential barrier in this case. Hence, **XIII** cannot be classified as a stable conformer, although the potential energy surface has a true local minimum at this conformation. The low barrier of 0.06 kJ/mol also influences the reaction paths of **XIII** to a large extent. In contrast to the other local minima, **XIII** can access only three distinct reaction paths that lead to saddle points.

Another feature that deserves special attention is the orientation of the groups $C = O$ and $O — H$. This orientation is characterized by a $H — O — C = O$ torsion angle around 180° in 10 of the 14 local minima; in **VI**, **VII**, **VIII**, and **XI**, however, the absolute value of the $H — O — C = O$ torsion angle is around 40° (see Discussion section).

The stability of the H-bonded local minima of ε-aminohexanoic acid is judged quite differently by the usual criteria for hydrogen bonds. Ordering according to SCF energy gives the stability ranking

$$I > II > III > IV > V > VI > VII > VIII > IX > X > XI > XII \gg XIII > XIV;$$

inclusion of vibrational zero-point energies changes this ranking slightly to

$$I > II > III > IV > V > VI > VII > VIII > IX > XII > XI > X > XIII \gg XIV.$$

Different rankings are obtained if bond lengths and electron densities are considered: The $O — H$ bond length (with values for **VI**, **VII**, **VIII**, and **XI** reduced by the mean difference between *syn* and *anti* orientation of the groups $C = O$ and $O — H$) yields the ranking

X > II ≈ VII > I > XI > IV ≈ XII > XIV > III

$$≈ V ≈ XIII > IX ≈ VIII > VI,$$

the electron density in the critical point that characterizes the N—H hydrogen bond [13] gives the ranking

X > II > VII > I > XII > XI > IV > XIV > V > IX > III > VIII > VI > XIII,

and the N—H distance gives the ranking

X > II > I > XII > VII > IV > XIV > XI > V > IX > III > VIII > XIII ≫ VI.

The N—H bond order [14] yields another different ranking, namely,

VII > XI > I ≈ X > II ≈ V > IX > VIII

$$≈ XII > XIII > III > VI > IV > XIV.$$

The height of the lowest potential barrier, which determines kinetic stability, gives yet another ranking:

II ≫ I ≫ IV > III ≈ X > V > IX > VII

$$≈ XI > XIV ≈ VI > XII > VIII > XIII.$$

Discussion

Figure 2 displays a comparison of characteristic data of the N · · · H—O hydrogen bond in the series glycine, β-alanine, γ-aminobutyric acid, δ-aminopentanoic acid, and ε-aminohexanoic acid. This comparison shows that the strength of the hydrogen bond increases up to δ-aminopentanoic acid and then decreases from δ-aminopentanoic acid to ε-aminohexanoic acid. At the same time, the difference between the strongest and weakest conformers in terms of the displayed quantities is approximately constant for γ-aminobutyric acid, δ-aminopentanoic acid, and ε-aminohexanoic acid. From these facts, it is obvious that the eight-membered ring, which is formed in δ-aminopentanoic acid, is the one with minimal constraints. Both γ-aminobutyric acid and ε-aminohexanoic acid cannot form more stable hydrogen bonds, but due to opposite reasons: In γ-aminobutyric acid, the carbon chain is too short to allow closer contact between the two functional groups, and in ε-aminohexanoic acid, the chain is already too long. As a consequence, most H-bonded conformers of ε-aminohexanoic acid have rather distorted geometries with one main chain torsion angle around ±120°, which in the absence of intramolecular interactions is a typical value for transition states. Also, most H-bonded conformers contain repulsive interactions between hydrogen atoms with H—H distances as low as 85% of the sum of the van der Waals radii (cf. Table I).

   The length of the carbon chain in ε-aminohexanoic acid is, however, just long enough to allow nine-membered rings with the —COOH group in the distorted *syn* orientation in **VI, VII, VIII,** and **XI**. This is a remarkable contrast to glycine, β-alanine, γ-aminobutyric acid, and δ-aminopentanoic acid, for which, as already
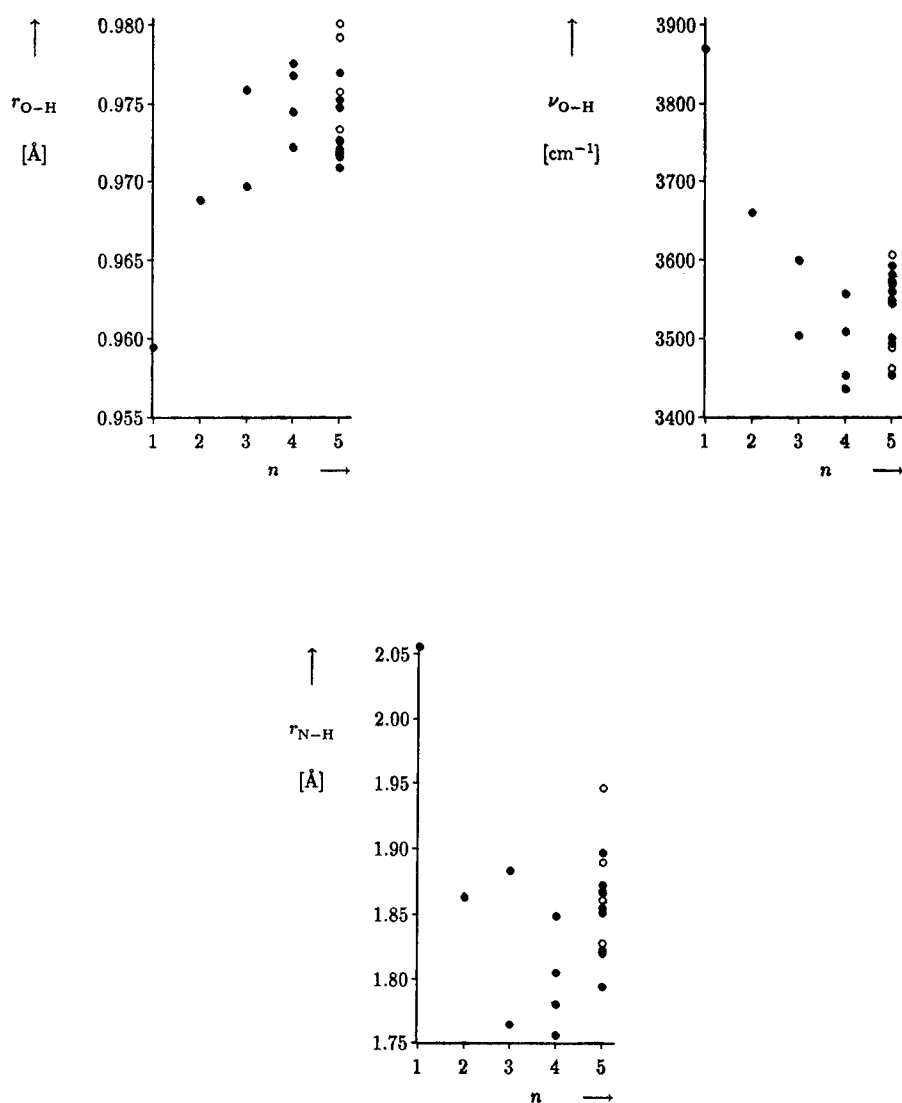
Figure 2.   O — H bond length and harmonic vibration frequency and N — H distance of all H_2N — $(CH_2)_n$ — COOH conformers with an intramolecular N $\cdots$ H — O hydrogen bond: ($\bullet$) *anti-periplanar* orientation of C=O and O — H; (O) *syn-clinal* orientation of C=O and O — H.

mentioned, the — COOH group occurs exclusively in *anti-periplanar* orientation in all H-bonded conformers. The *syn* orientation in **VI**, **VII**, **VIII**, and **XI** does not lead to lower energies, because it is too distorted from the H — O — C=O $\approx$ 0° orientation that is the most stable one in the absence of any interactions that involve the — COOH group. The large amount of distortion is also evident from

the pattern of hydrogen bond preserving reaction: **VI** and **VIII** both exhibit only one H bond preserving reaction path, **VII** and **XI** have two such reaction paths, whereas all ε-aminohexanoic acid conformers with $H \longrightarrow O \longrightarrow C = O \approx 180°$ take part in three or more such reaction paths.

If the trends, which can be gathered from ε-aminohexanoic acid and its homologs with fewer carbon atoms, are extrapolated further, one is led to the expectation that similar 10-membered rings will allow hydrogen-bonded conformers in which the terminal groups are present in their energetically most favorable orientation. In such conformers, hence, the stabilization due to the hydrogen bond would not be canceled by steric hinderings or unfavorable orientations, as is the case in all ω-amino acids up to and including ε-aminohexanoic acid. This expectation should also be generalizable to other organic compounds with intramolecular hydrogen bonds, because the specific interaction $N \cdot \cdot \cdot H \longrightarrow O$ in the ω-amino acids has been found *not* to be the dominating structural feature [15] in this class of compounds. These considerations, therefore, might be an explanation for the fact that 10-membered hydrogen-bonded rings occur so often in peptides and proteins.

Acknowledgments

Bibliography

[1] S. Vishveshwara and J. A. Pople, J. Am. Chem. Soc. **99**, 2422 (1977).
[2] H. L. Sellers and L. Schäfer, J. Am. Chem. Soc. **100**, 7728 (1978).
[3] M. Ramek, Int. J. Quantum Chem., Quantum Biol. Symp. **17**, 45 (1990).
[4] M. Ramek, V. K. W. Cheng, R. F. Frey, S. Q. Newton, and L. Schäfer, J. Mol. Struct. (Theochem) **235**, 1 (1991).
[5] J. H. Jensen and M. S. Gordon, J. Am. Chem. Soc. **113**, 7917 (1991).
[6] M. Ramek and V. K. W. Cheng, Int. J. Quantum Chem., Quantum Biol. Symp. **19**, 15 (1992).
[7] M. Ramek, J. Mol. Struct. (Theochem) **208**, 301 (1990).
[8] M. Ramek and M. Flock, Amino Acids (in press).
[9] M. Ramek, Struct. Chem., accepted.
[10] C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).
[11] R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).
[12] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, Shujun Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, Jr., J. Comp. Chem. **14**, 1347 (1993).
[13] R. W. F. Bader, *Atoms in Molecules* (Clarendon, Oxford, 1990), p. 19.
[14] I. Mayer, Chem. Phys. Lett. **97**, 270 (1983).
[15] M. Ramek, M. Flock, A.-M. Kelterer, and V. K. W. Cheng, J. Mol. Struct. (Theochem) **276**, 61 (1992).

Hydrogen Bonding of Nucleotide Base Pairs: Application of the PM3 Method

TRICIA N. LIVELY, MARCUS W. JUREMA,
AND GEORGE C. SHIELDS

Department of Chemistry, Lake Forest College, Lake Forest, Illinois 60045

Abstract

The ability of the PM3 semiempirical quantum mechanical method to reproduce hydrogen bonding in nucleotide base pairs was assessed. Results of PM3 calculations on the nucleotides 2'-deoxyadenosine 5'-monophosphate (pdA), 2'-deoxyguanosine 5'-monophosphate (pdG), 2'-deoxycytidine 5'-monophosphate (pdC), and 2'-deoxythymidine 5'-monophosphate (pdT) and the base pairs pdA–pdT, pdG–pdC, and pdG(*syn*)–pdC are presented and discussed. The PM3 method is the first of the parameterized NDDO quantum mechanical models with any ability to reproduce hydrogen bonding between nucleotide base pairs. Intermolecular hydrogen bond lengths between nucleotides displaying Watson–Crick base pairing are 0.1–0.2 Å less than experimental results. Nucleotide bond distances, bond angles, and torsion angles about the glycosyl bond (X), the $C_{4'}$—$C_{5'}$ bond (γ), and the $C_{5'}$—$O_{5'}$ bond (β) agree with experimental results. There are many possible conformations of nucleotides. PM3 calculations reveal that many of the most stable conformations are stabilized by intramolecular C—H— — —O hydrogen bonds. These interactions disrupt the usual sugar puckering. The stacking interactions of a dT–pdA duplex are examined at different levels of gradient optimization. The intramolecular hydrogen bonds found in the nucleotide base pairs disappear in the duplex, as a result of the additional constraints on the phosphate group when part of a DNA backbone. Sugar puckering is reproduced by the PM3 method for the four bases in the dT–pdA duplex. PM3 underestimates the attractive stacking interactions of base pairs in a B-DNA helical conformation. The performance of the PM3 method implemented in SPARTAN is contrasted with that implemented in MOPAC. At present, accurate *ab initio* calculations are too time-consuming to be of practical use, and molecular mechanics methods cannot be used to determine quantum mechanical properties such as reaction-path calculations, transition-state structures, and activation energies. The PM3 method should be used with extreme caution for examination of small DNA systems. Future parameterizations of semiempirical methods should incorporate base stacking interactions into the parameterization data set to enhance the ability of these methods. © 1994 John Wiley & Sons, Inc.

Introduction

Quantum mechanical methods have been applied to the study of purine and pyrimidine bases and nucleosides [1a–l] as well as to hydrogen-bonded dimers of these species [2a–r]. The molecular structures of purines, pyrimidines, nucleosides, and nucleotides and their intermolecular complexes are the subject of several monographs [3a–d]. Recently, it was demonstrated that the PM3 semiempirical quantum mechanical method can calculate intermolecular hydrogen bonding in small polar molecules [4a–e]. In this article, results of PM3 calculations on the nucleotides 2'-deoxyadenosine 5'-monophosphate (pdA), 2'-deoxyguanosine 5'-monophosphate (pdG), 2'-deoxycytidine 5'-monophosphate (pdC) and 2'-

deoxythymidine 5'-monophosphate (pdT) and the base pairs pdA–pdT, pdG–pdC, and pdG(*syn*)–pdC are presented and discussed. In addition, the effect of stacking DNA bases was assessed by PM3 calculations of a dT–pdT single strand and a dT–pdA duplex. This research was performed to assess the ability of the PM3 quantum mechanical method to reproduce the hydrogen-bonded structures and energies of nucleotide base pairs and to assess the ability of the method to model base stacking.

Method

The nucleotides pdA, pdG, pdC, and pdT were built with Chem 3D Plus software (Cambridge Scientific Computing, Inc., Cambridge, MA) on a Macintosh II computer (Apple Computer, Inc., Cupertino, CA). The nucleotides were built in the *anti* orientation about the glycosyl C_1·sugar–N base linkage. In addition, a monomer of pdG in the *syn* orientation was built. All nucleotide structures were designed as neutral species, by placing hydrogens on the appropriate phosphate oxygens. The five monomers were fully geometry-optimized with the PM3 method [5] using MOPAC 5.0 and 6.0 software [6], with the keywords PRECISE and NOMM. Base pairs were constructed from optimized monomers with initial hydrogen bond angles of 180° and hydrogen bond distances of 1.7 Å and minimized in the same fashion. All structures were characterized as stationary points and true minima using the keyword FORCE. Hydrogen bond energies were calculated by comparing the heats of formation of the base pairs with the 1SCF heats of formation for the monomers when "frozen" in their base pair-optimized geometry. The "frozen" monomers were then allowed to fully relax to calculate a second minima for each nucleotide. In addition, after examining the results of these calculations, several optimized monomer structures were perturbed slightly, removing or adding intramolecular hydrogen bonds. These structures were fully optimized to assess differences in local minima. Calculations were performed using the Cray Y-MP supercomputer (Pittsburgh Supercomputer Center) and a VAX 4300 (Lake Forest College).

In addition, several charged and neutral pdA–pdT dimers were calculated with MOPAC. The lowest energy charged and neutral structures were used as input for PM3 calculations within SPARTAN (Wavefunction, Irvine, CA). SPARTAN was also used to calculate the energies and structures of stacked nucleotide bases and base pairs. The starting structure was obtained from the middle portion of an A-tract DNA dodecamer crystal structure [7], with methyl groups placed in the C_5· position to replace the 5' phosphate and a hydrogen replacing the phosphate attached to the O_3· of the sugar. Hydrogens were added to the crystal structure to obtain a reasonable starting structure. All SPARTAN calculations were performed using an INDY workstation (Silicon Graphics, Mountain View, CA).

Results

The geometry-optimized structures of the pdA–pdT, pdG–pdC, and pdG(*syn*)–pdC neutral base pairs are presented in Figure 1. All hydrogens within 2 Å of another atom are illustrated by dashed lines. The hydrogen bonds responsible for holding the base pairs together are labeled with the appropriate hydrogen bond

A

B

C
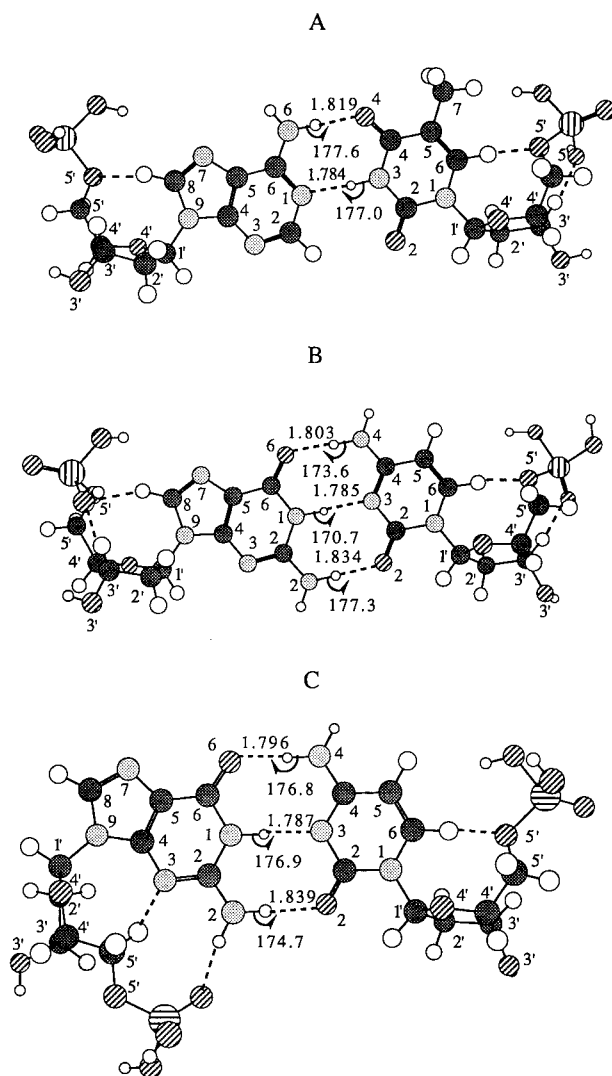
Figure 1. PM3-calculated structures of the nucleotide base pairs (A) pdA–pdT, (B) pdG–pdC, and (C) pdG(*syn*)–pdC. (●) C; (○) H; (◉) N; (⊘) O; (⊕) P.

distance and bond angle in the figure. Intermolecular hydrogen bond lengths range from 1.78 to 1.84 Å and hydrogen bond angles are between 170 and 178° for the three nucleotide base pairs. Torsion angles most relevant to nucleotide structure [3c] and sugar-puckering values are reported in Table I. In addition, the top third of the table reports the minimized heats of formation for each optimized monomer, the middle third lists the heats of formation for each monomer when frozen in the dimer configuration (the 1SCF calculations), and the bottom of the table contains

TABLE I.   Torsion angles, sugar-puckering values, and heats of formation for PM3 calculated nucleotide dimers and monomers.

	Monomers (optimized)[a]				
Parameter	pdA	pdT	pdG	pdC	pdGsyn
χ (O4'—C1'—N1—C2)		−122.8		−128.9	
(O4'—C1'—N9—C4)	−133.2		−126.6		55.8
γ (O5'—C5'—C4'—C3')	36.6	44.7	42.5	48.4	−61.7
β (P—O5'—C5'—C4')	−132.4	−113.6	−112.0	−115.3	169.2
P	−24.0	43.3	63.5	64.0	64.6
V_{max}	22.4	−7.15	−12.09	−5.02	−8.85
ΔH_f	−276.0	−408.06	−323.96	−344.67	−326.58

	Dimers					
Parameter	pdA	pdT	pdG	pdC	pdGsyn	pdC (in Gsyn bp)
χ (O4'—C1'—N1—C2)		−122.0		−132.6		−125.4
(O4'—C1'—N9—C4)	−134.3		−137.7		62.6	
γ (O5'—C5'—C4'—C3')	49.2	47.7	47.8	50.5	−66.3	53.0
β (P—O5'—C5'—C4')	−160.3	−113.9	−112.4	−101.6	166.0	−156.3
P	−19.6	48.3	b	60.9	55.3	48.3
V_{max}	15.92	−6.01	b	−5.35	−12.30	−3.96
$\Delta H_f(1\text{SCF})$[c]	−277.5	−407.2	−322.0	−342.2	−325.0	−343.6

	Monomers (fully relaxed from dimer conformation)					
χ (O4'—C1'—N1—C2)		−122.6		−135.2		−131.9
(O4'—C1'—N9—C4)	−134.7		−137.6		62.5	
γ (O5'—C5'—C4'—C3')	48.5	46.2	47.0	52.0	−66.2	54.4
β (P—O5'—C5'—C4')	−157.7	−113.9	−112.5	−101.3	165.8	−157.4
P	−26.4	53.5	−60.1	69.6	54.3	73.1
V_{max}	13.17	−5.88	9.23	−4.89	−11.82	−2.75
ΔH_f	−278.0	−408.07	−324.2	−343.6	−326.7	−346.3

[a] The same initial pdC was used to build the pdG–pdC and pdG(syn)–pdC dimers.

[b] $\nu_2 = 0$; P is undefined.

[c] Heat of formation for the monomer when "frozen" in the dimer conformation.

the heats of formation for each monomer after full minimization starting from the dimer configuration.

The PM3 heats of formation for the pdA–pdT, pdG–pdC, and pdG(syn)–pdC base pairs are −691.42, −679.79, and −683.56 kcal mol^{-1}, respectively. The heats of association, i.e., the difference between the base-pair heats of formation and the sum of the individual nucleotides "frozen" in the base-pair minima structure, are −6.72, −15.59, and −14.96 kcal mol^{-1} for the pdA–pdT, pdG–pdC, and pdG(syn)–pdC base pairs, respectively. Dividing the heats of association by the number of hydrogen bonds allows assignment of −3.36 kcal mol^{-1} per hydrogen bond between

pdA and pdT, -5.20 kcal mol^{-1} per hydrogen bond between pdG and pdC, and -4.99 kcal mol^{-1} per hydrogen bond between pdG(*syn*) and pdC.

Four pdA local minima were found, separated by 2.7 kcal mol^{-1}. Four pdC minima were located, all within 3 kcal mol^{-1} of each other. For pdG, the pdG(*syn*) monomer was 2.5 kcal mol^{-1} lower in energy than the lowest *anti* pdG molecule, which, in turn, was 0.25 kcal mol^{-1} lower in energy than any other local minima. No significant structural differences were found in these low lying minima of each individual nucleotide, except that one pdA and one pdC nucleotide lacked an intramolecular hydrogen bond.

Results of PM3 calculations on the stacked dT–pdA duplex, with an overall charge of -2, are displayed in Figure 2 as a function of gradient optimization termination conditions. The top structure in the figure comes from a DNA dodecamer crystal structure, with hydrogens added within SPARTANS builder module [7]. This structure was the input for PM3 calculations within SPARTAN. The middle structure results from geometry optimization with the gradient tolerance set at 0.001 au and the energy gradient set to 0.05 kcal mol^{-1}. The bottom structure results from geometry optimization under the conditions of 0.00005 au gradient tolerance and 0.001 kcal mol^{-1} energy tolerance. Table II contains torsion angles, sugar-puckering values, and heats of formation for the PM3-calculated dT–dpA duplex. Table III compares the sugar puckering for the pdA–pdT dimer, the input dT–pdA duplex from the crystal structure, and the dT–pdA duplex obtained from PM3 geometry optimization (**B** in Fig. 2). Figure 3 illustrates how PM3 unstacks one of the dT–dpT strands in the absence of hydrogen bonding.

Discussion

Individual Base Pairs

Intermolecular Hydrogen Bonds. The hydrogen bond distances between the pdA–pdT, pdG–pdC, and pdG(*syn*)–pdC base pairs are labeled in Figure 1. Comparing the distances between the donor and acceptor atoms for the pdA–pdT base pair, the PM3 calculated values of 2.827 Å for the A:N$_6$–T:O$_4$ acceptor/donor and 2.819 Å for the T:N$_3$–A:N$_1$ acceptor/donor atoms are slightly less than the experimental values of 2.950 and 2.820 Å from X-ray crystal structures [3c]. Leach and Kollman recently calculated intermolecular hydrogen bond distances and bond angles for the guanine–cytosine and adenine–thymine base pairs with the PM3 method [2r]. Their calculated values of 2.83 and 2.82 Å and 175.4 and 175.8° for the N$_6$ — O$_4$ and N$_3$ — N$_1$ acceptor/donor pairs differ from ours only slightly with respect to hydrogen-bond angles. This illustrates that the PM3 method does not treat hydrogen bonding of bases differently from hydrogen bonding of nucleosides or nucleotides. The PM3 hydrogen-bond energy of -3.36 kcal mol^{-1} is lower than the gas-phase hydrogen-bond energy of -6.5 kcal mol^{-1} determined by temperature-dependent field ionization mass spectrometry (TD-FIMS) [8]. It was previously shown that the PM3 method underestimates hydrogen-bond energies by as much as several kcal mol^{-1} and underestimates hydrogen-bond distances by 0.1–0.2 Å for small polar molecules [4d]. Leach and Kollman reported lower PM3 association energies for

LIVELY, JUREMA, AND SHIELDS
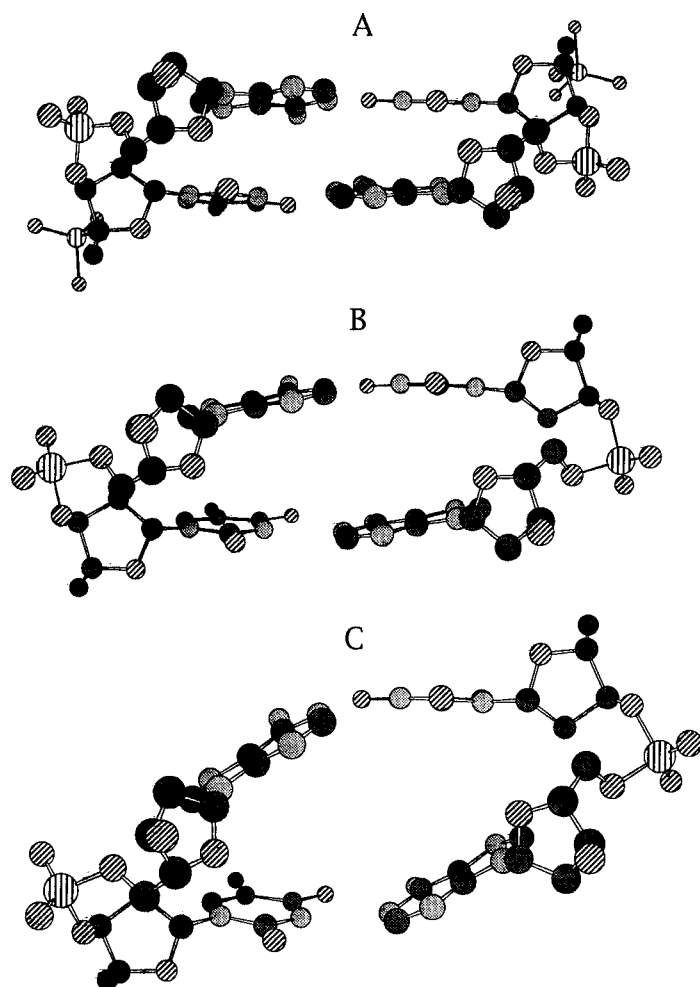
Figure 2. (A) Input geometry of a dT–pdA duplex taken from reference seven. (B) PM3 geometry after calculations in SPARTAN with an rms G = 0.0003 and energy changes of less than 0.05 kcal mol^{-1}. (C) PM3 geometry after further optimization in SPARTAN that reduced rmsG to less than 0.0000 with energy changing by less than 0.002 kcal mol^{-1}. (●) C; (○) H; (◎) N; (⊘) O; (◍) P.

base pairing, a consequence of their method of subtracting the sum of the individual fully minimized bases from that of the minimized base pair [2r].

For the pdG–pdC base pair, the PM3 donor/acceptor distances of 2.811, 2.806, and 2.844 Å are slightly shorter than the experimental values [3c] of 2.910, 2.850, and 2.860 Å for the $C:N_4$–$G:O_6$, $G:N_1$–$C:N_3$, and $G:N_2$–$C:O_2$ donor/acceptor distances, respectively. Compared to previous PM3 results [2r] on the base pair G–C, only the hydrogen bond angles differ by 2–5° for the nucleotide base pairs. The experimental TD-FIMS pdG–pdC gas-phase hydrogen bond energy is −7 kcal mol^{-1}

TABLE II. Torsion angles and sugar-puckering values for the dT–pdA duplex calculated with the PM3 Hamiltonian within SPARTAN.

Parameter	dT	pdA	dT	pdA
Input				
χ (O4'—C1'—N1—C2)	−97.8		−98.0	
(O4'—C1'—N9—C4)		−97.8		−98.0
γ (O5'—C5'—C4'—C3')	36.4	36.6	36.3	36.3
β (P—O5'—C5'—C4')	−146.0	−146.0	−146.0	−146.0
P	11.78	11.75	11.73	11.83
V_{max}	−35.65	−35.75	−35.64	−35.66
Output				
χ (O4'—C1'—N1—C2)	−94.3		−96.3	
(O4'—C1'—N9—C4)		−112.4		−112.0
γ (O5'—C5'—C4'—C3')	a	32.3	a	33.0
β (P—O5'—C5'—C4')	a	−152.4	a	−153.2
P	−4.19	19.1	−11.64	12.2
V_{max}	−27.6	−20.1	−27.1	−18.3

a Unable to do because phosphate group was removed.

[8] within 2 kcal mol^{-1} of the PM3 calculated value of −5.20 kcal mol^{-1}. The PM3 donor/acceptor distances for the pdG(syn)–pdC base pair are similar to the values for the pdG–pdC base pair, 2.807, 2.816, and 2.845 Å for the $C:N_4$–$G:O_6$, $G:N_1$–$C:N_3$, and $G:N_2$–$C:O_2$ atom pairs, respectively. The PM3 hydrogen bond energy is −4.99 kcal mol^{-1}.

TABLE III. Pseudorotation angles of the sugars.

Base	V0	V1	V2	V3	V4	P	V_{max}
Dimer output conformation							
A	10.2	−15.9	15.0	−9.7	0.0	−19.6	15.9
T	3.0	0.9	−4.0	6.0	−5.7	48.3	6.0
Stacked base pairs—input							
pdT	−4.2	24.9	−34.9	33.2	−18.3	11.78	−35.65
pdA	−4.3	25.0	−35.0	33.3	−18.4	11.75	−35.75
pdT	−4.3	24.9	−34.9	33.2	−18.3	11.73	−35.64
pdA	−4.2	24.9	−34.9	33.3	−18.3	11.83	−35.66
Stacked base pairs—output							
dT	−11.0	25.0	−27.5	22.3	−7.5	−4.19	−27.57
pdA	0	12.4	−19.0	19.8	−12.8	19.06	−20.10
dT	−14.0	26.0	−26.5	19.4	−3.8	−11.64	−27.06
pdA	−2.2	13.0	−17.9	17.4	−9.7	12.19	−18.31

tan $P = [(V4 + V1) − (V3 + V0)]/[2 \times V2 \times (\sin 36 + \sin 72)]$. $V_{max} = V2/\cos P$.

Intramolecular Hydrogen Bonds. The nucleotides that compose the base pairs in Figure 1(A) and (B) are in the *anti* position about the glycosyl $C_1 '$ — N torsion angle and in the $+sc$ orientation about the $C_4 '$ — $C_5 '$ bond. This conformation is stabilized by the C — H — — — $O_5 '$ hydrogen bonds of lengths 1.830, 1.851, 1.837, and 1.836 Å for pdA, pdT, pdG, and pdC, respectively. The hydrogen bond angles for these bonds are 164.0, 173.8, 159.9, and 173.9°, respectively. The PM3 calculations are consistent with X-ray crystal structures and deuterium-exchange NMR experiments, which show that the C_8 purine and C_6 pyrimidine protons are considerably acidic and participate in hydrogen bonding [3c]. The PM3 *anti* base pairs also have $C_3 '$ — H — — — O_P intramolecular hydrogen bond lengths of 1.874, 1.871, and 1.870 Å and angles of 165.8, 166.6, and 165.9° for the pdT, pdG, and pdC nucleotides, respectively. The main structural difference between the optimized dimers is that for the nucleotides in the *anti* orientation the pdA that is half of the pdA–pdT dimer and the pdC that is half of the pdG(syn)–pdC dimer do not have the $C_3 '$ — H — — — O_P intramolecular hydrogen bond. This is reflected by the β torsion angle, which is close to $-110°$ for the *anti* nucleotides that have the $C_3 '$ — H — — — O_P intramolecular bond, but changes to approximately $-160°$ for the pdA and pdC nucleotides that do not have this close contact.

For the pdG(syn)–pdC base pair displayed in Figure 1(C), the *anti, +sc,* conformer of pdC is stabilized by the C_6 — $O_5 '$ hydrogen bond 1.829 Å long. The pdG(syn) nucleotide has apparent hydrogen bonds of 1.888 Å, 154.5° and 1.815 Å, 169.0° for the $C_5 '$ — H — — — N_3 and N_2 — H — — — O_P parameters, respectively. In solution and in crystal structures, pdG nucleotides prefer the *syn* form, and previous calculations have implicated van der Waals and electrostatic attractions between the amino group in the 2 position and the 5' phosphate [3c].

Dimer Bond Distances and Bond Angles. The PM3 calculated bond distances and bond angles for nucleotides involved in intermolecular base pairing compare well with averaged data compiled from X-ray crystallographic studies [3c]. The standard deviation is 0.029 Å for bond lengths and 2.1° for bond angles for the nonhydrogen atoms in the three investigated base pairs. As shown in Table I, the torsion angle about the glycosyl bond, X, is *anti* ($-ac$) for all but the pdG(syn, sc) nucleotide. Rotation about the exocyclic $C_4 '$ — $C_5 '$ bond allows $O_5 '$ to assume three main conformations relative to the furanose ($\gamma = +sc, ap,$ or $-sc$). The three ranges are not uniformly populated, and for nucleotides, torsion angles γ and X fall in the $+sc$ and *anti* ranges, whereas for the pdG(syn) nucleotide, γ and X are in the *ap* and *syn* ranges [3c]. The PM3 results, outlined in Table I, show that all of the *anti* nucleotides are indeed $+sc$ with respect to the torsion angle γ, whereas the $C_5 '$ — H — — — N_3 hydrogen bond in the pdG(syn) nucleotide results in a torsion angle of $-66.3°$ for γ, falling in the $-sc$ range.

Sugar-puckering modes of the furanose ring are described by the pseudorotation cycle. In nucleotide structures, two ranges of pseudorotation angles are preferred: $C_3 '$-*endo* at $-18° < P < 36°$ (North) and $C_2 '$-*endo* at $130° < P < 200°$ (South) [9]. Although puckering at the $C_3 '$ and $C_2 '$ atoms is most commonly observed, the furanose ring is disordered in some crystal structures [10]. Interconversion between $C_3 '$-*endo* and $C_2 '$-*endo* puckering modes is extremely rapid, requiring approximately

5 kcal mol^{-1} activation energy [3c]. The sugar puckers calculated by PM3 are on the North side of the pseudorotation cycle, with $C_{4'}$-*exo* being the major puckering mode for all but the pdA nucleotide. For pdA, the furanose is twisted with major $C_{2'}$-*exo* and minor $C_{1'}$-*endo* puckering. A recent comparison of different computational methods for the conformational analysis of ring systems shows that AM1 and PM3 calculations optimized dideoxyribose to a near planar configuration [10]. Our results with ν_{max} values no higher than 16° for pdA in the dpA–pdT dimer (Table I) shows the inability of PM3 to accurately model the furanose ring in nucleotides.

Crystal structures of nucleotides show that the torsion angle β, which defines rotation about the $C_{5'}$—$O_{5'}$ bond, is limited mainly to the *ap* range with some structures in the *ac* range [3c]. The nucleotides pdA, pdG(*syn*), and pdC (from the pdG *syn* dimer) are in the normal *ap* range, whereas pdT, pdG, and pdC are in the *ac* range. Only the structures in the *ac* range form the $C_{3'}$—H— — —O_P intramolecular hydrogen bond.

Monomer Bond Distances and Bond Angles. The lowest-energy PM3 monomer nucleotide structures have bond lengths and bond angles consistent with averaged crystallographic data [3c]. The standard deviations are 0.25 Å for bond lengths and 2.1° for bond angles. All the intramolecular hydrogen bonds found in the dimer calculations also exist in the minimized monomers. Table I reveals that rotations about the glycosyl bond (χ), the $C_{4'}$—$C_{5'}$ bond (γ), the $C_{5'}$—$O_{5'}$ bond (β), and preferred sugar puckerings for the monomers deviate only slightly from the dimer structures. Nucleotides are flexible molecules, and the PM3 calculations illustrate that many local minima are possible. Slight adjustments to the models followed by full minimization produced a pdA molecule with the $C_{3'}$—H— — —O_P intramolecular interaction and a pdC molecule without the $C_{3'}$—H— — —O_P close contact. Our lowest-energy structure for the pdA nucleotide monomer does not show this close contact, but another local minima with the intramolecular bond is 2.1 kcal mol^{-1} higher in energy.

SPARTAN Base Pairs. The implementation of PM3 within SPARTAN differs from that in MOPAC. Standard tolerances for termination of geometry optimization are much tighter in SPARTAN than in MOPAC. In addition, there is an error in the calculation of the hydrogen bond energy for charged nucleotide base pairs in MOPAC that is partly rectified in SPARTAN. Optimization of the anionic pdA–pdT base pair (charge = −2) in MOPAC, followed by the 1SCF calculations on the individual nucleotides (each with a charge of −1), gives positive rather than negative association energies. Positive association energies on the order of hundreds of kcals mol^{-1} with MOPAC are reduced to approximately 20 kcal mol^{-1} with SPARTAN. The problem is not completely corrected in SPARTAN, but the error is certainly reduced.

Comparison of the charged and uncharged pdA–pdT dimers calculated with the PM3 Hamiltonian within SPARTAN reveal that the overall structure remains fairly similar between the charged and uncharged species. The only angle that changes in a significant way is for the rotation about the $C_{5'}$—$O_{5'}$ bond (β). The value for β changes from −161° to −128° for pdA and from −105° to −114° for pdT. This

change is easily accommodated as the phosphate group is very flexible. All the intramolecular hydrogen bonds present in the neutral dimer remain in the doubly charged dimer.

Comparison of the SPARTAN neutral pdA–pdT dimer with the MOPAC dimer reveals that the two minima are virtually identical. The SPARTAN values for χ, γ, β, V0, V1, V2, V3, and V4 of $-134.6°$, $47.3°$, $-160.8°$, $10.3°$, $-14.8°$, $13.3°$, $-7.9°$, and $-1.4°$, respectively, for pdA are within $2°$ of those for the MOPAC dimer reported in Tables I and III. Similarly, the SPARTAN values for χ, γ, β, V0, V1, V2, V3, and V4 of $-122.8°$, $46.0°$, $-113.8°$, $3.5°$, $0.2°$, $-3.3°$, $5.5°$, and $-5.7°$, respectively, for pdT are very close to those values reported for the MOPAC dimer. The heat of formation for the SPARTAN pdA–pdT uncharged dimer is lowered by 0.07 kcal mol^{-1} from the MOPAC calculation (with PRECISE).

Stacked Base Pair

Figure 2 and Table II contain the results of PM3 calculations on the stacked dT–pdA system. Table II analyzes structure **B** in Figure 2. Table II reveals that the torsion angles and overall sugar-puckering values are very similar for each nucleotide in the input structure **A** obtained from X-ray crystallography [7]. Comparing the input with the PM3 output shows that the crystal fragment relaxes in this gas-phase calculation, with the biggest change in χ for pdA. The $15°$ change in χ rotates the adenine further away from the plane of the sugar, still well within the *anti* range. Table III shows how stacking DNA bases influences the sugar puckering within PM3. The input structure for the stacked dT–pdA duplex is quite puckered, whereas the output from the PM3 calculation is less puckered but still much more puckered than is the pdA–pdT dimer. The phosphate groups in DNA molecules lose much of the flexibility observed in a single base pair, and the intramolecular hydrogen bonds observed between the phosphate oxygens and the sugar C—H disappear in the dT–pdA duplex. The constraints of the phosphate group remove the intramolecular hydrogen bonds, and as a result, the sugar puckering is reproduced much better in the duplex structure.

Although the PM3 method reproduces intermolecular hydrogen bonds very well for a parameterized method, it is not as successful for the stacking interactions of DNA bases. The driving force for DNA helix formation is believed to come from stacking interactions of DNA bases rather than from hydrogen bonding between the base pairs [3c]. This is because in solution DNA bases are already hydrogen-bonded with water molecules, and pairing of the bases can only occur after breaking base–water hydrogen bonds. Structures **B** and **C** in Figure 2 show that geometry optimization is driving the bases apart. The heat of formation for **B** is -824 kcal mol^{-1}, the rms$G = 0.0003$, and the energy is changing by less than 0.05 kcal mol^{-1}. In **C**, the heat of formation is -830.758 kcal mol^{-1}, the rms$G$ is less than 0.0000, and the energy is changing by less than 0.002 kcal mol^{-1}. The hydrogen bonds between the base pairs are a stabilizing influence on the stacking disruption. Figure 3 shows that removing the hydrogen bonds allows the bases to quickly unfold (heat of formation $= -554.841$
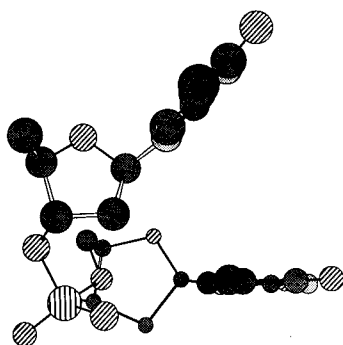
Figure 3. PM3 geometry-optimized structure of a dT–pdT single strand. (●) C; (○) H; (◉) N; (⊘) O; (◍) P.

after converging within the default gradient tolerances). The PM3-optimized dT–pdT structure unstacked the thymine bases and formed the $C-H---O_{5'}$ intramolecular hydrogen bond. This intramolecular interaction was not present in the input geometry. These calculations suggest that if a DNA helix of many base pairs were geometry-optimized by the PM3 method that the middle bases would hold together fairly well, but that the terminal bases would be subject to unstacking at high gradient tolerances.

Conclusions

The PM3 semiempirical quantum mechanical method is capable of reproducing the structures of nucleotides and nucleotide base pairs. Nucleotide bond distances, bond angles, torsion angles about the glycosyl bond (χ), torsion angles about the $C_{4'}-C_{5'}$ bond (γ), and torsion angles about the $C_{5'}-O_{5'}$ bond (β) agree with experimental results. Hydrogen bond energies are underestimated by 2–3 kcal mol^{-1}. Intermolecular hydrogen bond lengths between nucleotides displaying Watson–Crick base pairing are 0.1–0.2 Å less than experimental results. Sugar puckering is not reproduced by the PM3 method for nucleotide base pairs, although results do fall on the north side of the pseudorotation cycle. There are many possible conformations of nucleotides. PM3 calculations reveal that many of the most stable conformations are stabilized by intramolecular $C-H---O$ hydrogen bonds. Stacking of bases constrains the phosphate connecting the sugars in DNA, removing the intramolecular $C-H---O$ hydrogen bonds between the sugar and the phosphate, with the result that sugar puckering is now modeled quite well. PM3 stacking interactions are repulsive, rather than attractive, with the result that more stringent gradient tolerances tend to slowly unstack the dT–dpA duplex. It is quite likely that PM3 calculations on a bigger helix with less stringent gradient tolerances than are the norm in SPARTAN would adequately model the center portion of the helix. Future pa-

rameterizations of semiempirical methods should include base-stacking interactions in the parameterization set. Until that time, PM3 should be used with caution for modeling DNA chemistry.

Acknowledgments

Bibliography

[1] (a) J. Ladik, P. Otto, and W. Förner, Int. J. Quantum Chem., Quantum Biol. Symp. **10**, 73 (1983). (b) A. Buda and A. Sygula, J. Mol. Struct. (Theochem) **92**, 255 (1983). (c) M. Scanlan and I. H. Hillier, J. Am. Chem. Soc. **106**, 3737 (1984). (d) M. Aida, K. Yamane, and C. Nagata, Mutat. Res. **173**, 49 (1986). (e) Z. Latajka, W. B. Person, and K. Morokuma, J. Mol. Struct. (Theochem) **135**, 253 (1986). (f) U. Norinder, J. Mol. Struct. (Theochem) **151**, 259 (1987). (g) P. Hobza and C. Sandorfy, J. Am. Chem. Soc. **109**, 1302 (1987). (h) J. Kwiatkowski, R. Bartlett, and W. B. Person, J. Am. Chem. Soc. **110**, 2353 (1988). (i) H. Basch, D. R. Garmer, P. G. Jasien, M. Krauss, and W. J. Stevens, Chem. Phys. Lett. **163**, 514 (1989). (j) A. Leś, L. Adamowicz, and R. Bartlett, J. Phys. Chem. **93**, 4001 (1989). (k) M. Sabio, S. Topiol, and W. C. Lumma, Jr. J. Phys. Chem. **94**, 1366 (1990). (l) M. Szczesniak, J. Leszczynski, and W. B. Person, J. Am. Chem. Soc. **114**, 2731 (1992).

[2] (a) A. Imamura, H. Fujita, and C. Nagata, Bull. Chem. Soc. Jpn. **40**, 522 (1967). (b) R. Rein, P. Claverie, and M. Pollak, Int. J. Quantum Chem. **2**, 129 (1968). (c) L. Seprödi, G. Biczó, and J. Ladik, Int. J. Quantum Chem. **3**, 621 (1969). (d) E. Clementi, J. Mehl, and W. von Niessen, J. Chem. Phys. **54**, 508 (1971). (e) Z. G. Kudritskaya and V. I. Danilov, J. Theor. Biol. **59**, 303 (1976). (f) V. A. Kuprievich, Z. G. Kudritskaya, and V. E. Klymenko, Int. J. Quantum Chem. **16**, 71 (1979). (g) J. Marañon, O. M. Sorarrain, H. Grinberg, S. Lamdan, and C. H. Gaozza, J. Theor. Biol. **74**, 11 (1978). (h) H. Grinberg, A. L. Capparelli, A. Spina, J. Marañon, and O. M. Sorarrain, J. Phys. Chem. **85**, 2751 (1981). (i) J. Marañon and H. Grinberg, J. Mol. Struct. (Theochem) **88**, 283 (1982). (j) A. Sarai and M. Saito, Int. J. Quantum Chem. **25**, 527 (1984). (k) C. Giessner-Prettre, J. Biomol. Struct. Dyn. **2**, 233 (1984). (l) W. Förner, P. Otto, and J. Ladik, Chem. Phys. **86**, 49 (1984). (m) P. Cieplak, U. C. Singh, and P. A. Kollman, Int. J. Quantum Chem., Quantum Biol. Symp. **14**, 65 (1987). (n) P. G. Jasien and G. Fitzgerald, J. Chem. Phys. **93**, 2554 (1990). (o) J. Pranata, S. G. Wierschke, and W. L. Jorgensen, J. Am. Chem. Soc. **113**, 2810 (1991). (p) M. Aida, J. Comp. Chem. **9**, 362 (1988). (q) A.-O. Colson, B. Besler, D. M. Close, and M. D. Sevilla, J. Phys. Chem. **96**, 661 (1992). (r) A. R. Leach and P. A. Kollman, J. Am. Chem. Soc. **114**, 3675 (1992).

[3] (a) D. Voet and A. Rich, Prog. Nucleic Acid Res. Mol. Biol. **10**, 183 (1970). (b) R. Taylor and O. Kennard, J. Mol. Struct. **78**, 1 (1982). (c) W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984). (d) G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures* (Springer-Verlag, New York, 1991).

[4] (a) J. J. P. Stewart, J. Comp. Chem. **10**, 221 (1989). (b) I. Juranic, H. S. Rzepa, and M. Yi, J. Chem. Soc. Perkin Trans. **2**, 877 (1990). (c) H. S. Rzepa and M. Yi, J. Chem. Soc., Perkin Trans. **2**, 943 (1990). (d) M. W. Jurema and G. C. Shields, J. Comp. Chem. **14**, 89 (1993). (e) Y.-J. Zheng and K. M. Merz, J. Comp. Chem. **13**, 1151 (1992).

[5] J. J. P. Stewart, J. Comp. Chem. **10**, 209 (1989).

[6] J. J. P. Stewart, QCPE 455, available from Indiana University, Creative Arts Building 181, Bloom-
     ington, IN.
[7] A. D. DiGabrielle, M. R. Sanderson, and T. A. Steitz, Proc. Natl. Acad. Sci. U.S.A. **86**, 1816
     (1989).
[8] L. Sukhodub, Chem. Rev. **87**, 589 (1987).
[9] E. Westhof and M. Sundaralingam, in *Structure and Dynamics: Nucleic Acids and Proteins*[1],
     (E. Clementi and R. H. Sarma, Eds. (Adenine Press, New York (1983), pp. 135–147.
[10] D. M. Ferguson, I. R. Gould, W. A. Glauser, S. Schroeder, and P. A. Kollman, J. Comp. Chem.
     13, 525 (1992).

Molecular Dynamics Simulation of Crystal Water with X-Ray Constraints

J. RAUL GRIGERA

IFLYSIB (CONICET-UNLP), c.c. 565, 1900 La Plata, Argentina

TOMAS S. GRIGERA

Departamento de Fisica, UNLP, c.c. 67, 1900 La Plata, Argentina

EDUARDO I. HOWARD

IFLYSIB (CONICET-UNLP), c.c. 565, 1900 La Plata, Argentina

ALBERTO PODJARNY

UPR de Biologie Structurale, IBMC, 15 Rue Descartes, Strasbourg, France

Abstract

The study of water in macromolecular crystals is approached with a restrained molecular dynamics method that makes use of X-ray diffraction data, without the need of thermal B factors for the solvent. This method, called here solute-grid-restrained molecular dynamics (SGRMD), is applied to a test case of a simulated crystal of erythrol. The results are quite satisfactory, and it is concluded that the method can be useful to study real macromolecular crystals. © 1994 John Wiley & Sons, Inc.

Introduction

The problem of obtaining the contribution of the solvent to X-ray diffraction intensities remains a major task. For the case of ordered water molecules visible in a difference map, it is highly labor-intensive and dependent on a possible "human factor" in the interpretation of density peaks as water molecules. For the case of the contribution of disordered solvent, we can cite the work of Blake et al. [1], who used a uniform solvent density; that of Cheng and Schoenborn [2], who modeled the disordered solvent with closely spaced pseudoatoms; and that of Badger and Caspar [3], who used an iterative density modification procedure to obtain a "solvent density."

In this work, we proposed to obtain simultaneously a complete solvent model, both for ordered and disordered molecules, as well as the corresponding density. To do so, we devised an "experimentally biased" molecular dynamics simulation of a fully solvated crystal. The X-ray experimental data act as a perturbation on the modelization forces, in a way different from the one used in simulated annealing. Instead of calculating a contribution based on atomic positions, which require the knowledge of B-factors, an intermediate occupancy grid is used both for accumu-

lating water trajectories and for introducing forces dependent on electron density difference maps calculated from X-ray amplitudes.

Algorithm

All code was written based on the existing GROMOS MDX package (BIOMOS); permission for modification was granted by the owners (Profs. H. J. C. Berendsen and W. F. van Gunsteren). All additional routines were also written in FORTRAN. These additional routines concern the calculation of the solvent occupancy grid and of the perturbation force.

The Occupancy Grid

The method is based on the construction of a fine grid covering the solvent region of the simulation box. Each grid point will have associated a water occupancy. The first step consists of filling a box—of the same dimensions of the crystal unit cell—with the solute molecule, which has been already refined by standard methods, and randomly distributed water molecules. This system is run through dynamic steps, imposing position restraints on the solute and letting the water move following the prescriptions for molecular dynamics simulation. Water trajectories are kept and used to calculate the initial occupancy grid (Fig. 1, step 1). To measure the instantaneous occupancy, the solvent volume is divided in pixels (smaller than a water molecule) centered at grid points. Each pixel is considered to be occupied
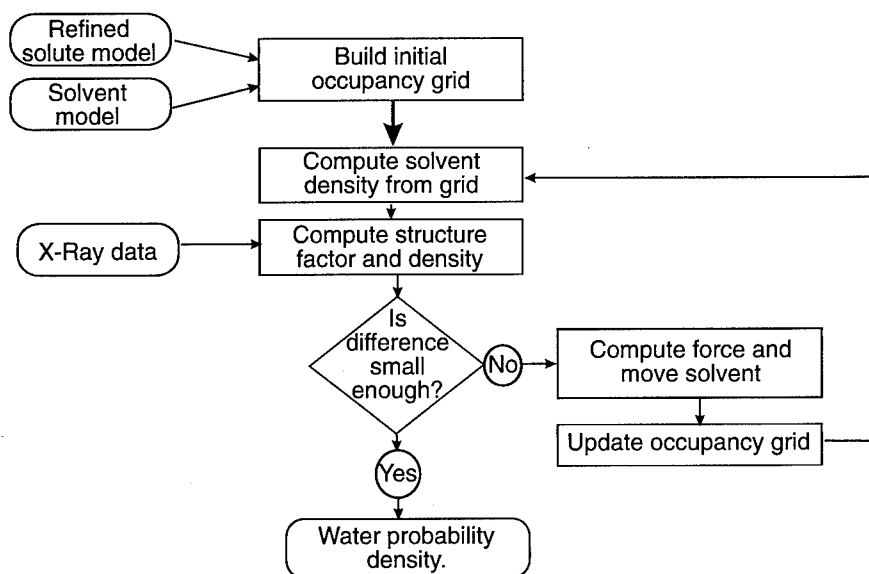

Figure 1.   Flowchart of the procedure.

by a water molecule if the center of the molecule lies inside it. This water molecule also contributes to neighboring pixels, the contribution diminishing with the distance. The instantaneous occupancy of the pixel is then associated to the grid point. The mean occupancy of a particular grid point at step n [$Oc(n)$] is defined as a weighted average of the instantaneous occupancies (Oc_{inst}). Since the system is evolving, old configurations must be gradually discarded and, also, the instantaneous occupancy has to be gradually introduced. To accomplish this without having to store old configurations, the occupancy in the nth step is computed as

$$Oc(n) = [Oc(n - 1)(L - Q) + Oc_{inst}Q]/L. \tag{1}$$

The ratio Q/L gives the weight applied to the incoming configuration. In our tests, $Q = 1$ and $L = 100$.

The Perturbation Force

The simulation proceeds starting from the previously calculated grid, applying to each water molecule an extra force, as described in what follows. Mean water occupancies at grid points yield an absolute solvent electron density, considering that a fully occupied site corresponds to the scattering of a water molecule (oxygen atom). As indicated before, the occupancy of a grid point for a molecule gives a contribution of the electron density in an environment that cover a number of grid points. The electron density is computed as

$$\rho(i,j,k) \propto \sum_{l=i-p}^{l+p} \sum_{m=j-p}^{j+p} \sum_{n=k-p}^{k+p} W(I_n)Oc(l,m,n), \tag{2}$$

where p is the maximum neighboring grid point that contributes to electron density and $W(Index)$ is the weight of the contribution of the I_n neighbor level.

This solvent electron density (Fig. 1, step 2) contributes then to the calculated structure factor F^{cal}. From the total calculated structure factor (the contribution of the solute molecule and the solvent grid) and the experimentally observed structural factor F^{obs}, a difference density map,

$$\Delta\rho = \mathscr{F}\{\|F^{obs}| - |F^{calc}\| \exp(i\phi^{calc})\}, \tag{3}$$

is obtained (Fig. 1, step 3). A positive value for the density difference in a point will indicate the lack of water molecules in such a point, and a negative value, an excess. This density difference map can be used to obtain a map for each grid point that will generate a force centered in the grid point, attracting or repelling water molecules, according with the needs to agree with the experimental data. This force is scaled to be of the same order of the modelization forces. This scale can be varied in order to weight more or less the crystallographic contribution. The difference density map is not an instantaneous picture of the solvent distribution but an average over a number of solvent configurations. It is updated fast enough to allow a convenient advance and slowly enough to avoid fast fluctuations. All the dynamics now evolve having the two contributions: the modeled forces and the grid-mediated

forces (Fig. 1, step 4). We can call the method the solute-grid-restrained molecular dynamics (SGRMD).

Testing

The method has been tested on pseudoexperimental data for erythrol with three types of hydration water. Five molecules "fixed" (with a thermal factor $B = 4$), five in the "second hydration shell" ($B = 16$), and 16 as a "bulk water" ($B = 80$). The water model corresponds to the SPC/E model [4] and erythrol has been modeled and simulated already [5]. Figure 2 shows the radial distribution of water molecules around the oxygen atoms during the last 10 ps of this simulation.

The test system has the advantage of being small—programs during the final debugging run fast—and that we know the results of the simulation, i.e., the behavior of the model. It has the disadvantage of being a rather artificial crystal. Some water molecules selected in the "first hydration shell" are known to be on sites of short lifetime due to the proximity of a hydrophobic corner of erythrol. In all these tests, the value of Q/L was 0.01, i.e., at every point, the occupancy is the average of 100 molecular dynamics steps.

Results

The original occupancy grid was obtained by averaging the solvent configurations from a molecular dynamics run of 600 steps of 0.002 ps without X-ray constraints.
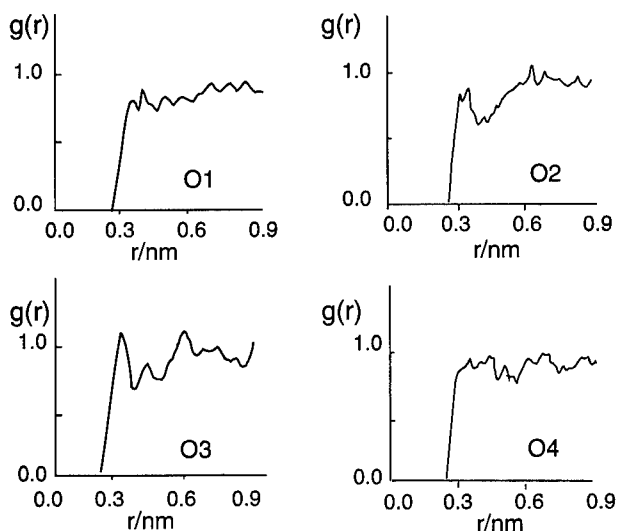
Figure 2. Radial distribution of water molecules around the oxygen atoms during the last 10 ps of the simulation of erythrol in water. Note the peaks at 0.3 nm (3) around O2 and O3, showing the existence of an hydrogen-bonded hydration shell. This peak is not present around O1 and O4, showing that they have a different electronic distribution that gives them a more hydrophobic character.

The X-ray restraints were then introduced. Two different resolutions were considered: 0.15 and 0.2 nm. At 0.15 nm resolution, four molecular dynamics runs of 500 steps of 0.002 ps each, were performed.

At 0.2 nm resolution, two sets of runs were done:

1. Four molecular dynamic runs of 500 steps of 0.002 ps each.
2. Twelve molecular dynamic runs of 500 steps of 0.002 ps each.

The results are shown (see Figs. 3–5) as the superposition of the original configuration on the simulated electron density. Some interesting features are observed:

- In the water occupancy map using forces from the 0.15 nm resolution data and averaging over 4 ps, many peaks appear, and all but one "fixed" water molecule are positioned in a peak or quite near to one. Water "channels" can be seen (Fig. 3).
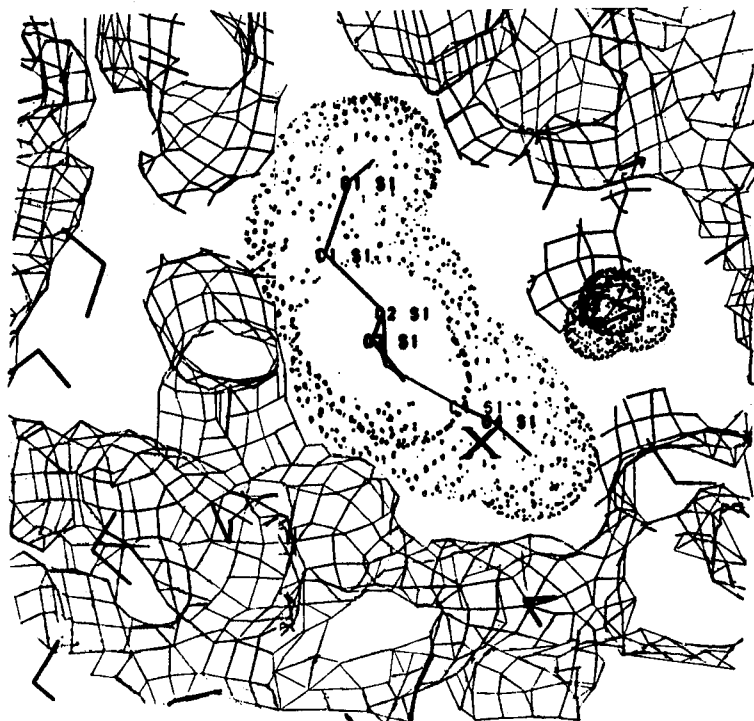
Figure 3.   This figure shows the erythrol molecule marked as S1 (Solute 1) and its surface. The oxygen O4 is marked by an X. It also shows the averaged solvent occupancy map (solid line lattice) after 4 ps of simulation using 0.15 nm resolution for X-ray forces, contoured at 0.4 of the maximum value, and the water molecules used for calculating the pseudo F_{obs}. One strongly diffracting water molecule near O4 is highlighted (dotted circles). Note that the neighborhood of this site is marked by the occupancy map. Note also the existence of water channels.
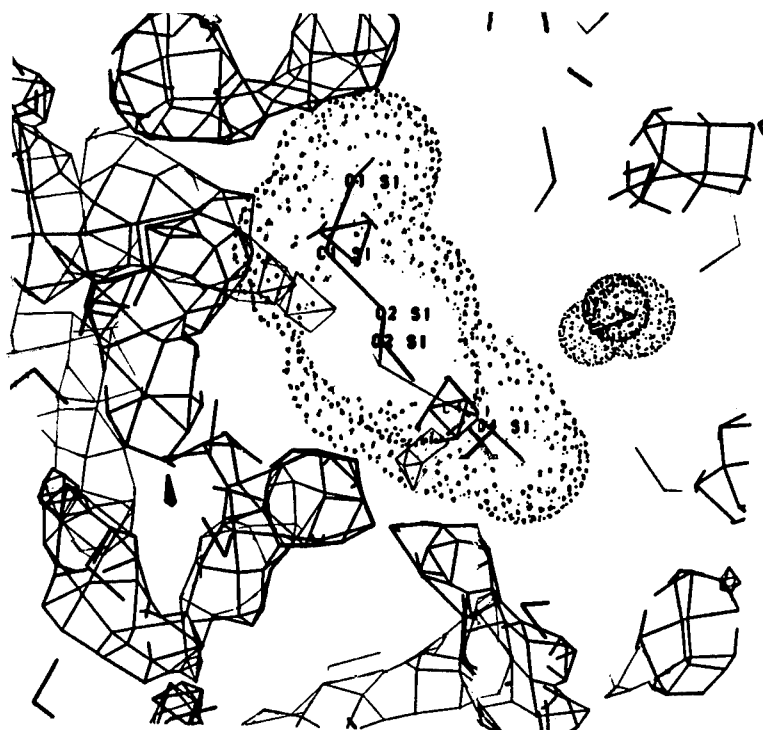
Figure 4. Same as Figure 3, but using 0.2 nm resolution X-ray data to calculate the X-ray forces. Note that the regions near O4 and O1 are much emptier. The map shows several peaks, which appear as small regions when contouring at 40% of the maximum value.

- The map corresponding to 0.2 nm resolution data and also averaging over 4 ps shows some remarkable differences. Around O4 of erythrol, we selected one of the "fixed" waters. As is known from simulation, the O4 is, for the model, a "hydrophobic oxygen" (as the radial distribution function shows in Fig. 2). Whereas in the previous map there was water near O4, in this case, the map grid is empty (Fig. 4). The model needs an empty space and the "created" X-ray forces need water, and, therefore, a "push–pull" process is produced, which here favors the model.

- A second map using 0.2 nm resolution data was obtained after averaging over 12 ps. The map is flatter and the water channels are less contrasted. It is seen that in this case the hydrophobic zone near O4 is occupied more (Fig. 5), meaning that the longer simulation time allows for the action of the crystallographic forces.

Conclusions

The algorithm seems to work properly. It tends toward an equilibrium where the crystallographic forces modify the occupancy grid, which shows the strongly diffracting water molecules.
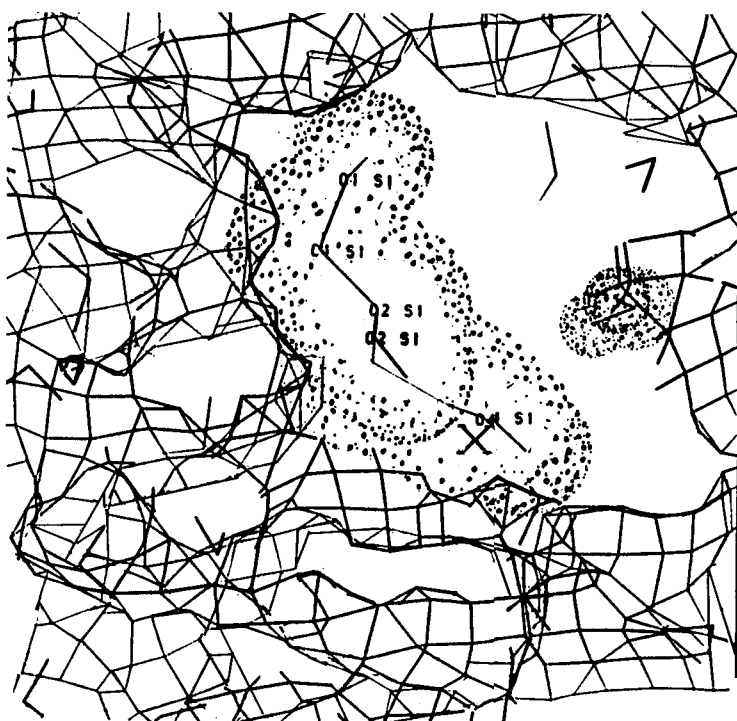
Figure 5. Same as Figure 4, but averaging over 12 ps instead of 4. Note that the map
tends to be flatter, as expected from a longer averaging time, and that the occupation
density now points toward the highlighted high occupation water molecule, meaning that
the longer simulation time is necessary for the action of the X-ray forces.

The analysis of hydration cannot be done only through the inspection of the
simulated density, however descriptive it might be. It should be complemented by
radial distribution functions and hydration lifetime studies on particular sites.

The results shown here are just a test of the algorithm. Therefore, the next step
will be the application to a real system. At present, we are using this method to
analyze experimental data obtained in one of our laboratories [6] of an RNA
tetradecamer.

Acknowledgments

J. R. G. is member of the Carrera del Investigador of CONICET and an E. I. Howard Fellow of the same institution.

Bibliography

[1] C. C. F. Blake, W. C. A. Pullford, and P. J. Artymiuk, *J. Mol. Biol.* **167**, 693 (1983).
[2] X. Cheng and B. P. Schoenborn, *Acta Crystogr.* **B46**, 195 (1990).
[3] J. Badger and D. L. D. Caspar, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 622 (1991).
[4] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.* **91**, 6169 (1987).
[5] E. I. Howard and J. R. Grigera, *J. Chem. Soc., Faraday Trans.* **88**, 437 (1992).
[6] A. C. Dock-Bregeon, B. Chevrier, A. D. Podjarny, J. E. Johonson, J. S. de Bear, G. R. Gough, P. T. Gilham, and D. Moras, *J. Mol. Biol.* **209**, 459 (1989).

Computational Studies of the Structures and Properties of Potential Antimalarial Compounds Based on the 1,2,4-Trioxane Ring Structure. I. Artemisinin-like Molecules

GÉRARD BERNARDINELLI AND CHARLES W. JEFFORD

Department of Organic Chemistry, University of Geneva, 1211 Geneva 4, Switzerland

DJORDJE MARIC

CSCS, Manno, Switzerland

COLIN THOMSON

Department of Chemistry, University of St. Andrews, St. Andrews, KY16 9ST, Scotland

JACQUES WEBER

Department of Physical Chemistry, University of Geneva, 1211 Geneva 4, Switzerland

Abstract

Artemisinin and related molecules are potential antimalarials that contain the 1,2,4-trioxane ring system. Several new derivatives have been synthesized and tested in Geneva, and this article presents the results of a systematic study of the structure of these molecules, both by the semiempirical PM3 method and using *ab initio* SCF methods. The results highlight the feasibility of full optimizations with 3-21G and 6-31G* basis sets for these large molecules. Molecular electrostatic potential (MEP) maps are evaluated and used in an attempt to identify the key features of the molecules that are necessary for their activity. There is good agreement between the PM3 and *ab initio* maps as to the qualitative predictions. © 1994 John Wiley & Sons, Inc.

Introduction

The worldwide emergence of resistance to the forms of malaria spread by the parasite *Plasmodium falciparum* has meant a large increase in the incidence of the disease and mortality from it, particularly in Africa and Asia [1]. Although there are several drugs that were formerly effective against this form of malaria, in particular, chloroquine, which accounts for 80–90% of all antimalarial drugs in current use [2], resistant strains are unaffected by this molecule and other chemically related nitrogen heterocycles. The search for new antimalarial drugs that are effective in this form of malaria thus has a very high priority in antimalarial drug design.

Fortunately, Chinese researchers have found a new lead compound, artemisinin (formerly Qinghaosu) (Fig. 1), (a), in extracts from herbs that have been used in China for thousands of years [3]. This compound is a sequiterpene containing the 1,2,4-trioxane ring structure (b), and a variety of derivatives of this molecule, namely,
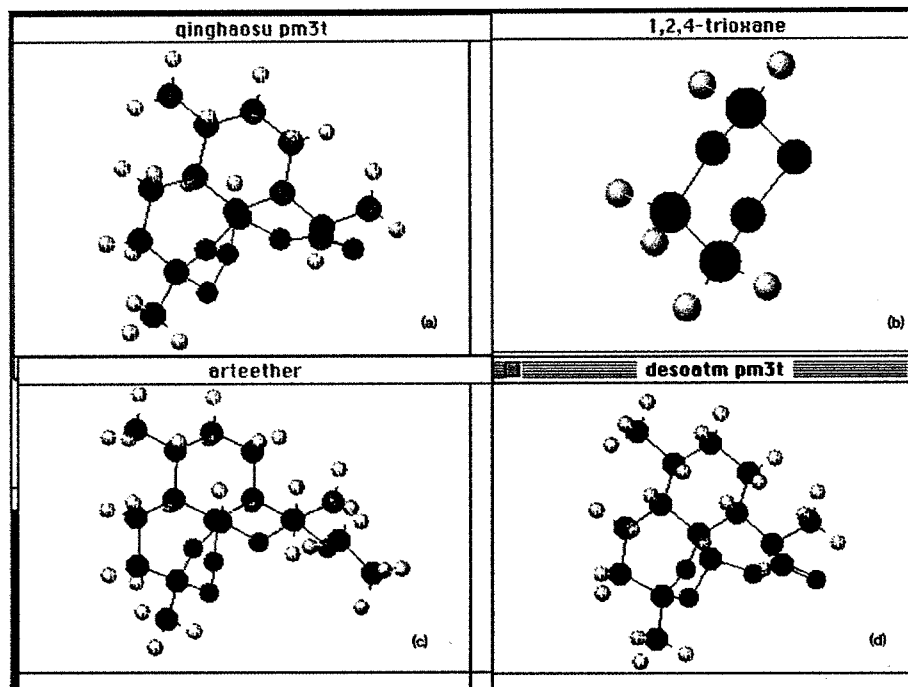
Figure 1.   (a) Structure of artemisinin; (b) 1,2,4-trioxane; (c) arteether; and (d) desoxyar-
temisinin.

arteether (c), and several other derivatives (d) have been chosen for clinical tri-
als [4].

However, all these compounds have disadvantages in terms of difficulties in ad-
ministration due to poor solubility or stability, and a more rational approach to
the development of better drugs of this type was clearly needed. One such approach
that we have adopted is to try to identify the unique features of these molecules
that are necessary for activity and to suggest new molecules for synthesis on the
basis of these features.

For several years, one of the authors (C. W. J.) has pursued this approach from
an experimental viewpoint [5], but in this article, we describe an alternative strategy,
which is to investigate the molecules developed in this experimental program by
theoretical techniques that we believe yields a more detailed rationalization of the
observed structure–activity data. We previously described some earlier work on (1)
[6], and the present article extends the calculations to several molecules that have
turned out to be more effective antimalarials [7], but which have not so far been
studied in detail by the methods of computational chemistry. A second article will
deal with a class of related molecules based on a *cis* fused cyclopenteno-1,2,4-
trioxane structure [8].

Background to the Development of the Relevant Molecules Related to Artemisinin

Several earlier studies have addressed the question of how much of (1) is necessary for antimalarial activity [3,7]. It is known that the endoperoxide link in (1) is necessary, because desoxyartemisinin (4), where only the ether bridge remains, is inactive [9].

A variety of studies on cyclic peroxides and tetrahydrobenzopyran-derived 1,2,4-trioxanes have shown that neither the peroxide function nor the 1,2,4-trioxane ring alone are sufficient to confer antimalarial activity. The rings in (1), labeled A, B, C, and D, were removed in earlier work by C. W. J. to give derivatives of varying activity. We refer to these compounds with the numbers given in the recent review by Jefford and co-workers [7] (Fig. 2). The antimalarial data is given in Table I.

Compound 25, in which ring D is removed, is active, but compound 19 containing the cyclopentane ring is more active. This shows that ring D is not necessary. The resulting systems may be called ABC structures and the X-ray structures of one of these molecules is known [10]. In (25), the trioxane ring is in the chair form, whereas it is in a distorted boat form in (1). Compounds based on the ACD ring structure have also been prepared [11]. In particular, compounds (22) and (23), which differ only by the CH_3O group being *endo* (22) or *exo* (23), are both active, but the *endo* isomer (22) is slightly more active.

The X-ray structures of these molecules [10] show the trioxane ring to be like that in (1), namely, a twisted boat. The inactive compound (26) has also been studied by X-ray diffraction. Derivatives in which the CH_3—group is replaced by Ph have also been prepared. These are also effective, more so than the CH_3—derivatives [10].

Further development of simpler molecules exploited the above observations. In particular, fusion of the cyclohexane A ring in a *cis* conformation with another ring system seemed like a potentially good structure.

Although the first series made, as fused naphtheno-1,2,4-trioxanes, were not active, a new series with *cis*-fused cyclohexeno derivatives were weakly active. However,

TABLE I. *In vitro* antimalarial activity of some artemisinin analogs against *P. falciparum* clones.

Compound	W2 clone (ng/mL)		D-6 clone (ng/mL)	
	IC_{50}	IC_{90}	IC_{50}	IC_{90}
1	1.1	—	2.2	—
19	2.0	3.3	2.3	30.4
25	6.2	25.8	28.7	39.8
22	1.8	3.9	16.5	11.5
23	9.7	16.9	75.4	8.75
26	Inactive		Inactive	

replacement of the cyclohexene ring with cyclopentene gave much more active derivatives, some of which approach (**1**) in activity [7].

Extensive synthetic work on these compounds has resulted in a great deal of structure activity (S/A) data [7]. The aim of the present article was to try to rationalize these data in terms of the molecular structure and properties of the molecules, and in this article, we focus on artemisinin [compound (**1**)] and the related the compounds (**19**), (**22**), (**23**), (**25**) and (**26**).

Computational Strategy

Theoretical Geometries. All the molecules that have antimalarial activity are relatively large and contain several rings with conformational flexibility. We previously showed [12] that reliable geometries of this type of molecule can be calculated by semiempirical SCF methods. We showed that the computed structure of (**1**) using the PM3 method was in excellent agreement [6] with the experimental X-ray structure [13] and this is the preferred method for calculations in this series of compounds [the AM1 method gives values of R(O—O) that are too long]. We found similar agreement for other molecules for which we have X-ray structures. Hence, our initial studies of the molecules in the present article were carried out at the PM3 level of SCF theory.

However, because we wish to use the molecular electrostatic potential (MEP) maps as a guide to the S/A relationships, we felt it necessary to evaluate the structures and wave functions computed by the *ab initio* SCF method using at least a split-valence (3-21G) basis set. For selected molecules, we also optimized the geometries with a 6-31G* basis set to check the reliability of the 3-21G predictions.

The geometry optimizations were carried out without any constraints, i.e., all $3N$-6 variables were optimized (Table II). This is probably not necessary for every variable, i.e., C—H bond lengths, but the lack of symmetry in most of these molecules makes it at least the consistent procedure for each one. It is also found that the optimizations proceeded more uniformly when all variables were optimized.

TABLE II. Computed values of E_{SCF} at the 3-21G and 6-31G* optimized geometries and no. variables optimized.

Molecule	Basis set		$3N$-6
	3-21G	6-31G*	
Artemisinin	−949.8037	−955.06882	120
Deoxyartemisinin	−875.4588	−880.3273	117
Jefford **19**	−915.3292	−920.3837	132
Jefford **22**	−761.2065	−765.407870	102
Jefford **23**	−761.2044	−765.40798	102
Jefford **26**	−761.1932	—	102
Pj26	−609.1052	—	—
Jefford **25**	−838.8552	−843.4582	120

Molecular Properties and S/A Correlations. Since the precise mechanism of action and relevant receptor is not known for these compounds, we attempted to rationalize the observed S/A data in two ways: We computed the MEP on molecular surfaces round each molecule and examined the features of the MEP as described below [14–16]. We also attempted to quantitatively compare various similarity indices [17], although this has so far not been as successful as hoped.

The Study of Model Compounds. In view of the expense of calculations as large as those described above, it is always sensible to investigate the structure of the basic molecular skeleton for the active drug molecule, but omitting all CH_3—or other groups, such as OH. These prototype molecules we call *pxx* and these have the basic ring structure of the active drugs. Optimization of these *pxx* compounds is relatively fast, and the main conformational features should be present in these compounds. We then add relevant CH_3—, OH, and C_6H_5—groups and reoptimize to get the full structure.

Computational Facilities and Software

This project has been carried out both at St. Andrews where the PM3 calculations were performed on a Tektronix CAChe molecular modeling system [18], and the *ab initio* calculations, on a two processor FPS-500 mini-supercomputer in St. Andrews, and in Geneva, where a Silicon Graphics Chrimson system was used for both SCF calculations and graphical display of the results. Some of the largest calculations were also carried out at the Swiss Supercomputer Centre at Manno on an NEC SX-3 supercomputer.

The programs used were mainly MOPAC-6 [19], GAUSSIAN 90 [20], and GAUSSIAN-92 [21], on our own machines, and to a lesser extent GAUSSIAN-92 on the SX3. In view of the number of basis functions needed (for most cases, at least 150), it was more efficient to carry out the SCF calculations using the DIRECT SCF method, and such calculations run very efficiently with GAUSSIAN-92.

Visualization of the results was done either on the CAChe system at St. Andrews or, more recently, using the MOLEKEL software developed in the Geneva laboratory by Flukiger [22]. MEP were computed from the Mullikan charges and displayed on a Connolly surface. We also examined the MEP computer with the Merz–Kollman procedure [23]. The MOLEKEL program is a sophisticated program for display of molecular structures and properties developed for Silicon Graphics workstations and has been extensively used in this study.

Results

Structural Aspects

It is convenient to divide the results into two sections that follow the historical development of these compounds by one of us. First, we compare artemisinin (**1**) with the active compounds containing the ACD ring and also with compounds (**19**) and (**25**). Second, we compare these active molecules with the inactive (**23**) and (**26**). The second article in this series will deal with the cyclopenteno derivatives.

BERNARDINELLI ET AL.

Comparison of Artemisinin (1), (19), (25), (22), (23) and (26). As pointed out in our earlier article, the calculated structure of (1) is in excellent agreement with experiment at the PM3 level [6]. The present *ab initio* results and the earlier results are given in Table II. It is particularly important to establish how well different levels of theory perform for this molecule, because of the expense of full optimizations for molecules containing so many atoms. The total energies for the 3-21G basis set are given in Table II and also for the 6.31G* basis set where these are available. Table III gives the calculated values of the ring parameters in artemisinin for the different methods. The agreement between the *ab initio* 3.21G and 6-31G* results is excellent, especially since the O1—O2 bond length is closer to experiment than in any of the semiempirical results. In artemisinin itself, the trioxane ring adopts a twist-boat conformation, and all the methods give the torsion angles quite well. It is interesting that both AM1 and PM3 give larger errors in the torsions than do the *ab initio* methods. The next set of active and inactive molecules studies are referred to as (j22), (j23), (j19), (j25), and (j26), and their formulas are given in Figures 2–5. X-ray data is available, and Tables IV–VI compare the results for (j22), (j23), and (j26).

There is again excellent agreement with the experimental data, and the ball-and-stick pictures are very similar to those in [10]. Bond lengths are in error by <0.02 Å in most cases, angles are within 2°, and the torsion angles in all are of the correct magnitude and sign, but individual angles may be in error by 2–3°, up to 7° for (j26). It is important to stress that with molecules as flexible as these better agreement with experiment is not expected. The basic structure of all these molecules is reproduced at the 3-21G level.

The results of the 6-31G* calculations for (j22) and (j23) are also given in Tables IV and V. The results are very similar to the 3-21G results, and this fact is in accordance with other studies by the authors. In general, it is not necessary to carry out 6-31G* optimizations. The structure obtained at the lower level of theory is

TABLE III.   Comparison of calculated and experimental values of the 1,2,4-trioxane ring parameters in artemisinin.

Parameter	AM1	PM3	ZINDO	Expt	6-31G*	3-21G
O1O2	1.289	1.544	1.240	1.478	1.390	1.462
O2C3	1.427	1.402	1.404	1.403	1.396	1.441
C3O4	1.427	1.428	1.402	1.437	1.408	1.436
O4C5	1.416	1.403	1.394	1.390	1.376	1.408
C5C6	1.537	1.555	1.499	1.529	1.532	1.529
O1O2C3	112.5	110.3	112.4	107.5	109.5	107.1
O2C3O4	103.6	104.8	106.7	107.3	107.8	107.3
C3O4C5	115.5	116.0	111.8	114.1	115.3	115.7
O4C5C6	113.5	115.2	114.1	113.3	112.3	112.1
O1O2C3C4	−77.7	−73.3	−76.9	—	−73.4	−74.6
O2C3O4C5	41.9	52.7	34.6	—	31.1	32.3
C3O4C5C6	11.5	2.8	21.1	—	27.4	28.3
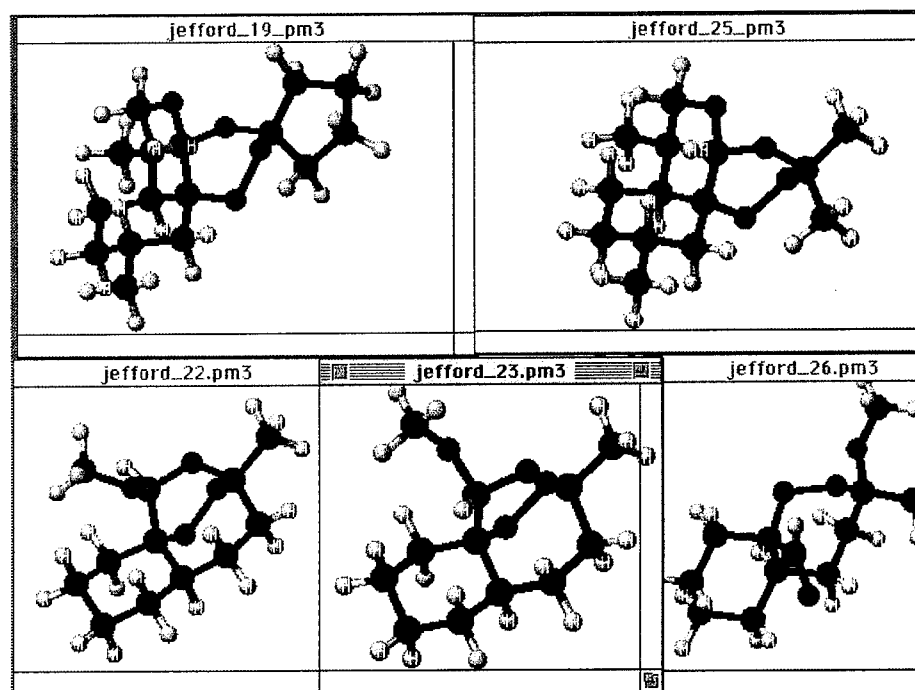
Figure 2. Structures of Jefford molecules: **j19**, **j25**, **j22**, **j23**, and **j26**.
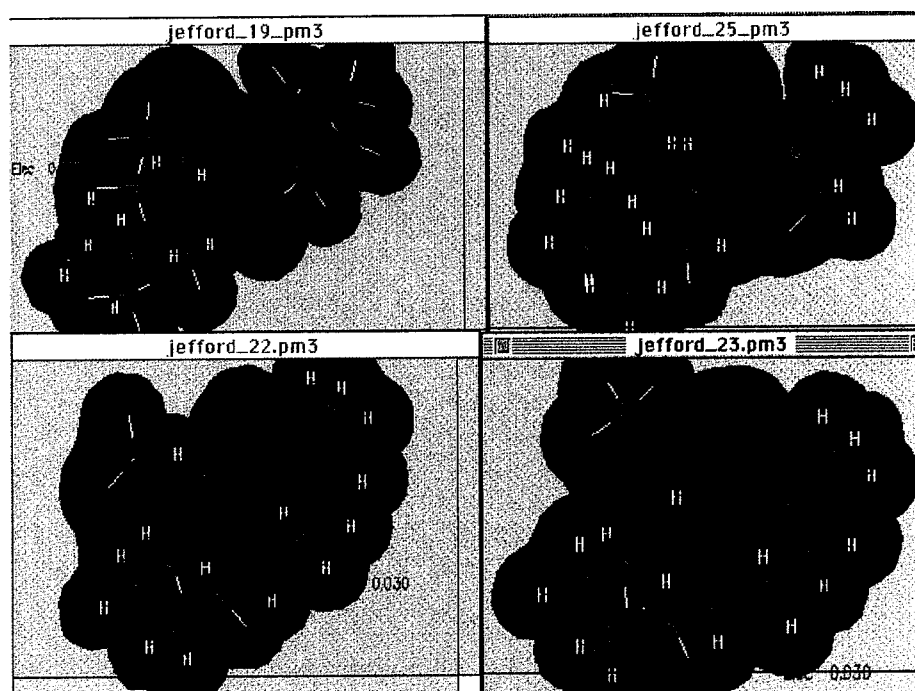
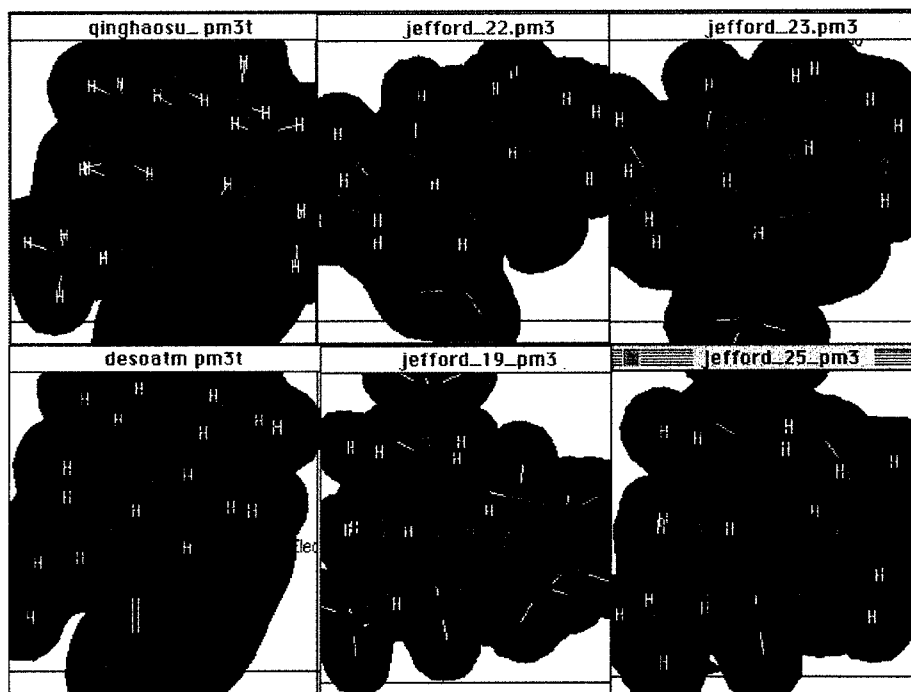Figure 3. MEP from the PM3 wave function for **j19**, **j15**, **j22**, and **j23**.

Figure 4.   Different view of the MEP for **j19**, **j25**, **j22**, and **j23**.
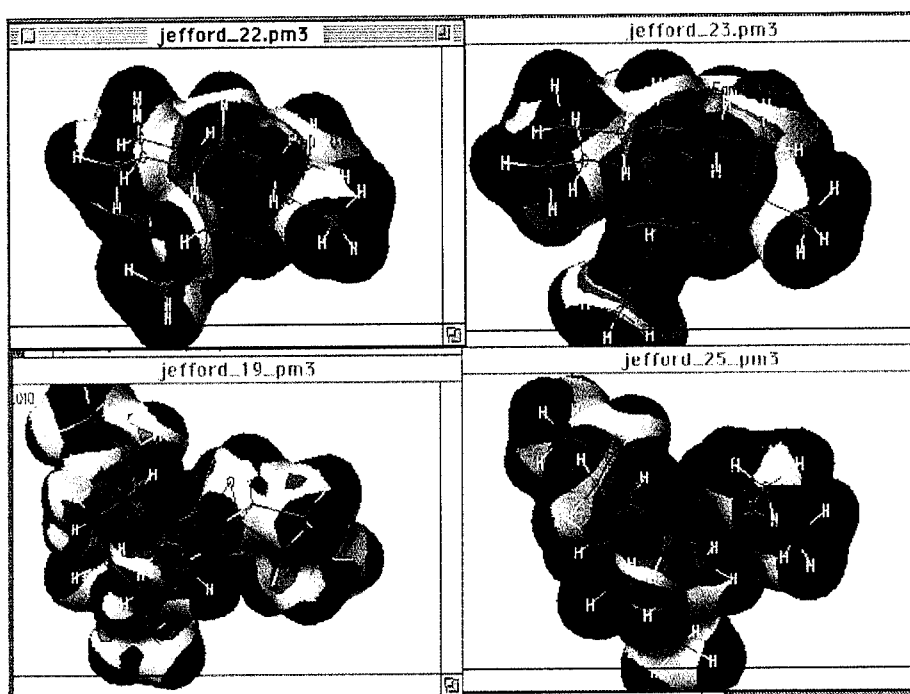

Figure 5.   MEP from the PM3 wave function on the electron density surface for **j19**, **j25**, **j22**, and **j23**.

TABLE IV.   Selected bond lengths (Å) and angles (deg) for **j22**.

	Expt	3-21G	Δ	6-31G*	Δ
C(1)—O(1)	1.417(5)	1.442	+0.025	1.398	−0.019
O(1)—O(2)	1.473(4)	1.465	−0.008	1.393	−0.08
O(2)—C(3)	1.473(5)	1.493	+0.02	1.440	−0.033
C(3)—C(2)	1.525(6)	1.529	+0.004	1.536	+0.011
C(2)—O(3)	1.427(4)	1.434	+0.006	1.401	−0.026
O(3)—C(1)	1.424(5)	1.433	+0.007	1.403	−0.021
O(4)—C(1)	—	3.558	—		—
O(4)—C(2)	1.391(5)	1.403	+0.012	1.371	−0.02
O(3)—C(1)—O(1)	107.7(2)	106.8	−0.9	107.5	−0.2
C(1)—O(1)—O(2)	110.1(2)	108.9	−1.3	111.1	+1.0
O(1)—O(2)—C(3)	112.5(3)	112.2	−0.3	113.8	+1.3
O(2)—C(3)—C(2)	106.2(3)	107.4	+1.2	106.8	+0.6
C(3)—C(2)—O(3)	112.5(2)	111.9	−0.6	111.7	−0.8
C(2)—O(3)—C(1)	113.3(3)	114.3	+1.0	114.4	+1.1
C(2)—O(4)—C(12)	113.0(3)	115.3	+2.3	115.3	+2.3
O(3)—C(1)—O(1)—O(2)	−71.5(3)	−73.2	−1.7	−71.5	0
C(1)—O(1)—O(2)—C(3)	39.9(3)	44.9	+5.0	42.6	+2.7
O(1)—O(2)—C(3)—C(2)	23.0(3)	+18.2	−4.8	19.3	−3.7
O(2)—C(3)—C(2)—O(3)	−61.5(3)	−58.8	−2.7	−57.5	−3
C(3)—C(2)—O(3)—C(1)	32.8(4)	33.0	+0.2	31.4	−1.4
C(2)—O(3)—C(1)—O(1)	32.1(3)	31.2	+0.9	30.2	−1.9
O(3)—C(2)—O(4)—C(12)	71.1(3)	58.0		64.4	−0.7
C(3)—C(2)—O(4)—C(12)	−166.7(3)	179.9		−172.9	−4
O(1)—O(2)—C(3)—C(8)	−102.6(3)	−105.4	−2.8	−104.5	−2
C(3)—C(8)—C(9)—C(10)	−41.9(5)	−44.1	−2.2	−41.1	−0
C(8)—C(9)—C(10)—C(1)	58.6(5)	62.8	4.2	57.5	−1
C(9)—C(10)—C(1)—O(1)	−95.0(4)	−97.2	2.2	−93.2	−2
C(10)—C(1)—O(1)—O(2)	50.8(3)	47.8	3.0	49.9	0.8

adequate for subsequent property calculations. The good agreement between the ring torsion angles and experiment is particularly gratifying, since it can be concluded that the ring structure of these molecules can be reproduced with some reliability. Of course, the other conformers of these rings will be close in energy, but we have not so far looked at these aspects of their structure.

It is also of interest to note that in the case of the inactive (**j26**), our computations of (**j26**) and (**pj26**), in which the CH_3 and CH_3O groups are replaced by H, show that these molecules have significant differences in structure, the (**j26**) results being closer to experiment. Clearly, these substituents *significantly* alter the 1,2,4-trioxane ring structure.

Turning now to (**j19**) and (**j25**), which are more closely related to artemisinin itself, the X-ray structure of (**j25**) has been determined previously [5]. Our overall results are again in good agreement, but when the two CH_3 groups are replaced by the cyclopentane ring, the trioxane ring conformation is altered (Table VII).

TABLE V.   Selected bond lengths (Å) for **j23**.

	Expt	3-21G	Δ	6-31G*	Δ
C(1)—O(1)	1.45(2)	1.436	−0.01	1.451	0
O(1)—O(2)	1.47(1)	1.462	−0.01	1.396	−0.07
O(2)—C(3)	1.48(1)	1.482	0	1.472	−0.01
C(3)—C(2)	1.50(1)	1.529	+0.03	1.574	+0.07
C(2)—O(3)	1.43(1)	1.428	0	1.440	+0.01
O(3)—C(1)	1.47(1)	1.435	−0.04	1.441	−0.03
O(4)—C(1)	—	3.176	—	3.548	—
O(4)—C(2)	1.40(1)	1.406	+0.01	1.430	+0.03
O(3)—C(1)—O(1)	106(1)	107.7	+1.7	108.0	+2.0
C(1)—O(1)—O(2)	109.8(7)	108.3	−1.5	108.8	−1.0
O(1)—O(2)—C(3)	111.6(7)	112.0	+0.4	110.6	−1.0
O(2)—C(3)—C(2)	109.7(7)	110.5	+0.8	108.8	−0.9
C(3)—C(2)—O(3)	113.7(8)	111.4	−2.3	112.9	−0.8
C(2)—O(3)—C(1)	113.7(9)	115.5	+1.8	110.2	−3.5
C(2)—O(4)—C(12)	113.9(9)	115.6	+1.7	110.2	−3.7
O(3)—C(1)—O(1)—O(2)	−73.4(8)	−73.0	−0.4	−79.3	−6.1
C(1)—O(1)—O(2)—C(3)	43(1)	46.6	+3.6	50.9	9.0
O(1)—O(2)—C(3)—C(2)	20(1)	14.6	−5.4	12.0	8.0
O(2)—C(3)—C(2)—O(3)	−57(1)	−53.9	−3	−56.3	−0.7
C(3)—C(2)—O(3)—C(1)	27(1)	29.4	+2.4	31.1	4.1
C(2)—O(3)—C(1)—O(1)	35(1)	31.9	−3.1		
O(3)—C(2)—O(4)—C(12)	−61(1)	−55.0	−6.0	−64.2	−3.0
C(3)—C(2)—O(4)—C(12)	173.0(9)	−178.1	−5.0	−172.7	−0.3
O(1)—O(2)—C(3)—C(8)	−103.3(9)	−106.7	−3.4	−110.8	−7.5
C(3)—C(8)—C(9)—C(10)	−40(1)	−41.1	−1.1	−42.2	−2.2
C(8)—C(9)—C(10)—C(1)	57(1)	61.0	+3.0	61.7	4.7
C(9)—C(10)—C(1)—O(1)	−95(1)	−97.4	−2.4	−94.9	+0.1
C(10)—C(1)—O(1)—O(2)	49(1)	47.7	−1.3	42.8	6.2

In (**j25**), the ring is quite close to the boat structure, but in (**j19**), the cyclopentane ring modifies this to a distorted twist boat structure. The differences in the torsion angles are as much as 18°. Table VIII gives the 1,2,4-trioxane ring parameters in the above compounds.

In summary, these results show that *ab initio* SCF computations with a split-valence basis set with *complete* geometry optimization can reproduce the experimental results for molecules of this complexity. It is important to emphasize that the starting geometry is **not** the X-ray structure, although it could have been. We just build a chemically reasonable structure with a suitable molecular modeling package, in our case, the CAChe system, and then optimize the structure as described above. The computed structures were of, course, verified to be minima using the force option. From these results, we are confident that even large molecular structures can be reliably computed using the direct SCF procedure on supercomputers such as the SX3. (Computer times for the 3-21G optimizations were up to few hours on the SX3 for the largest molecules.)

TABLE VI. Selected geometrical parameters for **Pj26** and **j26**.

	Expt (j26)	Pj26 3-21G	Δ	j26 3-21G
C(1)—O(1)	1.44(1)	1.451	+0.01	1.450
O(1)—O(2)	1.467(7)	1.464	+0.003	1.463
O(2)—C(3)	1.45(1)	1.462	+0.01	1.464
C(3)—C(2)	1.52(1)	1.515	−0.01	1.515
C(2)—O(3)	1.18(2)	1.208	+0.02	1.208
O(3)—C(1)	—	2.307	—	4.306
O(4)—C(1)	1.41(1)	—	—	1.408
O(4)—C(2)	—	—	—	4.744
O(3)—C(1)—O(1)	—	(37.2)		(39.5)
C(1)—O(1)—O(2)	111.5(6)	108.3	−3.5	110.8
O(1)—O(2)—C(3)	105.3(6)	104.1	−1.2	104.6
O(2)—C(3)—C(2)	106.6(7)	105.3	−1.3	105.5
C(3)—C(2)—O(3)	128(1)	125.8	−2.2	125.9
C(2)—O(3)—C(1)	—	(49.6)		(51.3)
C(2)—O(4)—C(12)	—	—		(101.5)
O(3)—C(1)—O(1)—O(2)	—	(−95.1)		(−95.9)
C(1)—O(1)—O(2)—C(3)	108.6(8)	110.8	+2.2	108.7
O(1)—O(2)—C(3)—C(2)	44.2(8)	44.1	−0.1	42.8
O(2)—C(3)—C(2)—O(3)	−138.1(9)	−143.8	−4.7	−142.3
C(3)—C(2)—O(3)—C(1)	—	(71.3)		(71.9)
C(2)—O(3)—C(1)—O(1)	—	(−40.8)		(27.5)
O(3)—C(2)—O(4)—C(12)	—	—		(−126.6)
C(3)—C(2)—O(4)—C(12)	—	—		(121.0)
O(1)—O(2)—C(3)—C(8)	−82.9(7)	−81.9	−1.0	−83.4
C(3)—C(8)—C(9)—C(10)	−72(1)	−72.1	−0.1	−70.2
C(8)—C(9)—C(10)—C(1)	85(1)	82.3	−3.7	84.6
C(9)—C(10)—C(1)—O(1)	−34(1)	−27.1	−7.0	−33.6
C(10)—C(1)—O(1)—O(2)	−51(1)	−58.2	−6.8	−51.5

Structure/Activity Correlations

The main aim of this project was to try to rationalize the observed s/a data on these compounds, which at first sight are not very similar. The 1,2,4-trioxane ring is, however, essential and our expectation was that the features of the MEP in this region would be informative in this respect.

We previously calculated the MEP on the Connolly surface for artemisinin and desoxyartemisinin from semiempirical SCF wave functions, using AM1 and PM3 Hamiltonians [6]. These results, using MEP computed from the wave function by the multipole expansion method, indicated significant differences in the MEP maps between the active and inactive molecules. The appearance of the maps depends, of course, on the direction of viewing, but it does seem that the negative potential region near the trioxane ring is narrower in the inactive molecule. This is actually seen very strikingly in the ZINDO MEP.

We believe, however, that more reliable MEP can be obtained by *ab initio* calculations of the MEP performed using the structures computed as described above.

BERNARDINELLI ET AL.

TABLE VII.   Selected bond lengths and angles in **j19** and **j25**.

	j19	**Exp**[a]	**j25**
C(1)—O(1)	1.434	1.412	1.449
O(1)—O(2)	1.461	1.477	1.464
O(2)—C(3)	1.480	1.472	1.487
C(3)—C(2)	1.541	1.528	1.532
C(2)—O(3)	1.416	1.403	1.418
O(3)—C(1)	1.439	1.447	1.450
O14—C5	1.419	—	1.419
O(3)—C(1)—O(1)	108.5	109.6	107.4
C(1)—O(1)—O(2)	105.8		107.5
O(1)—O(2)—C(3)	109.0		111.1
O(2)—C(3)—C(2)	108.1	105.6	108.4
C(3)—C(2)—O(3)	107.0	111.3	108.0
C(2)—O(3)—C(1)	115.1	116.0	116.2
C(2)—O(4)—C(12)	120.5	114.1	115.3
O(3)—C(1)—O(1)—O(2)	−48.7	61.3	−58.3
C(1)—O(1)—O(2)—C(3)	76.4	−71.9	67.3
O(1)—O(2)—C(3)—C(2)	−29.4	65.9	−11.2
O(2)—C(3)—C(2)—O(3)	−35.6	−55.1	−48.2
C(3)—C(2)—O(3)—C(1)	65.2	48.2	59.0
C(2)—O(3)—C(1)—O(1)	−20.3	−51.5	−4.4
O(3)--C(2)—O(4)—C(12)	154.0	176.4	171.0
C(3)—C(2)—O(4)—C(12)	34.3		49.8
O(1)—O(2)—C(3)—C(8)	—		—
C(3)—C(8)—C(9)—C(10)	—		—
C(8)—C(9)—C(10)—C(1)	—		—
C(9)—C(10)—C(1)—O(1)	—		—
C(10)—C(1)—O(1)—O(2)	—		—

[a] Values for **j25**.

We have the opportunity in this way to improve the wave function by using a larger basis set, if necessary.

The MEP can be computed *ab initio* in G92, and we have such studies underway, but we wish first to evaluate the use of MEP computed from Mulliken charges and also by the Merz–Kollman procedure in the S/A correlations. In the Merz–Kollman procedure, the charges are derived from a fit to the *ab initio*-computed potential on a grid round the molecule, and this may represent a more reliable method of computing the MEP than using semiempirical wave functions. Therefore, for the molecules described above, we proceeded as follows:

We first compute the optimized structure, and from the wave function at this geometry, a Mulliken population analysis and a Merz–Kollman analysis is carried out. The charges are then used to compute the MEP as a Connolly surface using the MOLEKEL program.

The results of the calculations are now described. There is no correlation between the values of the minimum in the MEP, but there are interesting variations in the

TABLE VIII. 124 trioxane ring parameters in different molecules

	Artemisinin	j22	j23	j19	j25
O1O2	1.462	1.465	1.462	1.461	1.460
O2C3	1.441	1.442	1.463	1.434	1.450
C3O4	1.436	1.433	1.435	1.430	1.439
O4C5	1.408	1.434	1.428	1.416	1.418
C5C6	1.529	1.529	1.529	1.541	1.540
O1C6	1.477	1.492	1.482	1.480	1.479
O1O2C3	107.1	108.9	108.3	105.8	106.1
O2C3O4	107.3	106.8	107.7	108.4	107.9
C3O4C5	115.7	114.3	115.5	115.2	115.5
O4C5C6	112.1	111.9	111.3	107.0	107.4
C5C6O1	111.6	107.4	110.5	108.1	108.2
O1O2C3O4	−74.6	−73.2	−73.0	−48.7	−49.7
O2C3O4C5	32.3	31.2	31.9	−20.3	−18.8
(O3O4C5C6	28.3	32.9	29.4	65.2	64.5
O4C5C6O)	−50.9	−58.8	−53.9	−35.6	−36.4
C5C6O1O2	10.0	18.2	14.6	−29.4	−28.2
C6O1O2C3	50.3	45.0	46.6	76.4	76.0

maps in this series. It should be noted that only (j26) and desoxyartemisinin are totally inactive, but because of inherent uncertainties in the biological data, it is reasonable to conclude that although artemisinin is the most active, (j19) and (j22) are of very similar activity, closely followed by (j25).

Figure 6 shows the MEP of these compounds. The details vary over the surface, but all the actual molecules have a region of negative potential of similar shape near the trioxane ring, but this region is **displaced** in the inactive compounds. This negative region is due to the peroxide linkage and is the most noticeable feature of the MEP, which is more similar in the cases of (j19), (j25), (j22), and (j23) than in artemisinin itself, because these four molecules have lost the $>C=O$ group that is present in artemisinin.

Discussion

The results obtained in this study shown both the strengths of weaknesses of the use of quantum chemistry calculations in attempting to rationalize S/A data as large molecules. First, we are able to reliably compute the structures of these large molecules with quite large basis sets, and the structures are likely to be quite reliable for those molecules where there are no experimental data on the molecules.

However, the wave functions calculated with this size basis set are probably accurate enough so that the computed MEP are reasonable approximations to the "exact MEP" and should reflect the potential felt by whatever the molecule interacts with *in vivo*. Unfortunately, although the active molecules have similar MEP round the essential trioxane ring, it has not proved possible to quantitatively compare
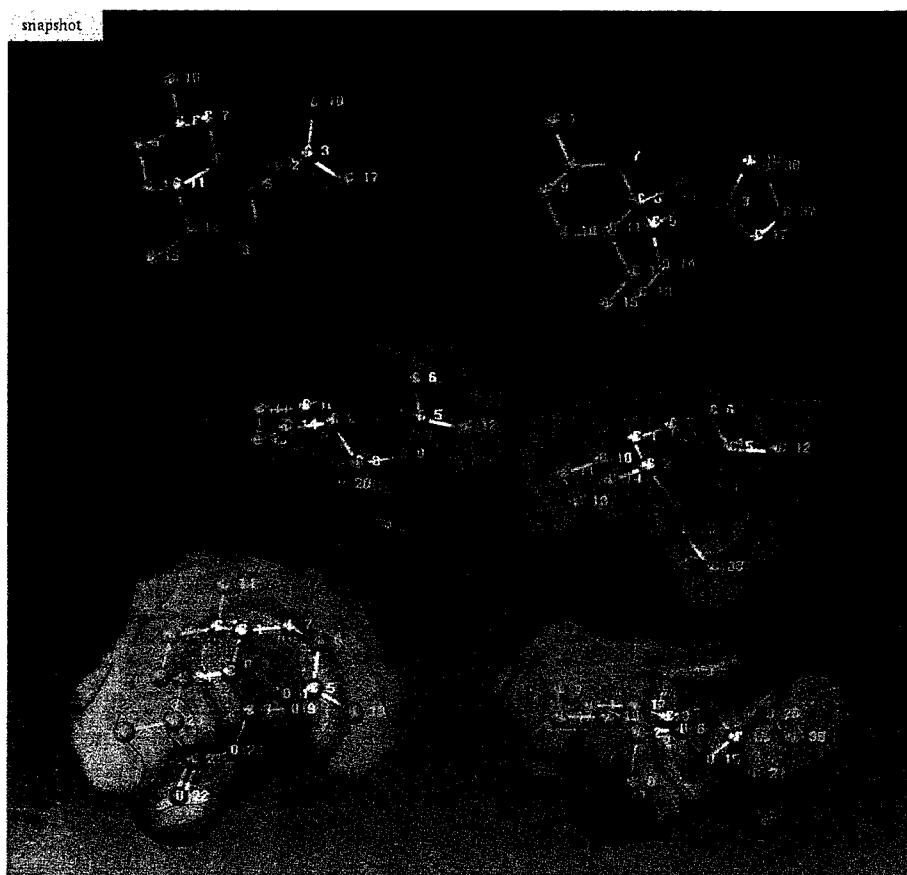
Figure 6.   MEP computed from the *ab initio* 6-31G* wave function using the MOLEKEL
program: j19, j25, j22, j23, artemisinin, and desoxyartemisinin.

these to our satisfaction, and, hence, to use this type of approach to predict new
and more effective molecules on the basis of their MEP. Partly, this is because
quantitative similarity indices cannot be evaluated with enough precision. However,
we are confident that any new active molecules should have MEP that are similar
to those found for artemisinin and the new Jefford molecules referred to above.

A further difficulty in this study is the lack of information on the detailed mechanism
of action of these compounds. Jefford et al. [10] postulated that the active molecule
interacts with haem to undergo electron transfer to the trioxane ring, resulting in ring
opening to give a radical anion. We earlier showed that addition of an electron to 1,2,4-
trioxane results in ring opening when we carry out optimization of the anion at the PM3
level. We confirmed this in *ab initio* calculations using the 6-31+G basis set; thus, this
mechanism may be a plausible one. We hope that we may carry out further calculations
on the interactions of the model antimalarials with a haem model in the near future.

Nevertheless, we believe that this study has resulted in some qualitatively useful information on the MEP of these molecules that should be of use in the design of new and active analogs of these interesting molecules.

Acknowledgments

Bibliography

[1] S. C. Oaks, Jr., V. S. Mitchell, G. W. Pearson, and C. C. J. Carpenter, Eds. *Malaria, Obstacles and Opportunities* (National Academy Press, Washington, DC, 1991); J. Maurice and A. M. Pearce, Ed. *Tropical Disease Research,* Eighth Programme Report (World Health Organization, Geneva, 1987); W. Peters, Br. Med. Bull. **38**, 187 (1982); W. Peters, Parasitology **90**, 705 (1985).

[2] *Practical Chemotherapy of Malaria,* WHO Tech. Rep. Ser. No 805 (WHO, Geneva, 1990).

[3] A. R. Butler and Y. Wu, Chem Soc. Rev. 85 (1992); D. L. Klayman, Science **228**, 1049 (1985).

[4] T. T. Hien and N. J. White, Lancet **341**, 603 (1993); G.-S. Ding, Intl. J. Exp. Clin. Chemother. **1**, 9 (1988).

[5] C. W. Jefford et al., Stud. Org. Chem. **31**, 113 (1986).

[6] C. Thomson, M. Corey, and M. Zerner, Int. J. Quantum Chem. Quantum Biol. **18**, 231 (1991).

[7] C. W. Jefford, D. Mista, J.-C. Rossier, P. Kamalaprija, U. Berger, J. Mareda, G. Bernardinelli, W. Peters, B. L. Robinson, W. K. Milhous, F. Zhang, D. K. Gosser, Jr., and S. R. Meshnick, in *Perspectives in Medicinal Chemistry,* B. Testa, W. Fuhrer, E. Kyburz, and R. Giger, Eds. (VCH, Amsterdam, 1992), Chap. 29.

[8] C. W. Jefford, C. Thomson, G. Bernadinelli, D. Maric, and J. Weber, Nov.

[9] A. Brossi, B. Venugopalan, L. Dominguez Gerpe, H. J. C. Yeh, J. L. Flippen-Anderson, P. Buchs, X. D. Los, W. Milhous, and W. Peters, J. Med. Chem. **31**, 645 (1988).

[10] C. W. Jefford, J. A. Velarde, G. Bernardinelli, D. H. Bray, D. C. Warhust, and W. K. Milhous, Nov.

[11] C. W. Jefford, J. Velarde, and G. Bernardinelli, Tetrahedron Lett. **30**, 4485 (1989).

[12] C. Thomson and P. Scano, J. Comput. Chem. **12**, 172 (1991).

[13] Institute of Biophysics, Academia Sinica, Sci. Sin. **22**, 1114 (1979).

[14] P. Politzer and D. G. Truhlar, Eds., *Chemical Applications of Atomic and Molecular Electrostatic Potentials* (Plenum, New York, 1981).

[15] C. Thomson, D. Higgins, and C. Edge, J. Mol. Graph. **6**, 171 (1988).

[16] D. Higgins, PhD Thesis (University of St. Andrews, St. Andrews, Scotland, 1988).

[17] Unpublished work.

[18] Tektronix CAChe System, Tektronix (1989).

[19] J. J. P. Stewart, Mopac Version 6.0 (U.S.A.F. Academy, Colorado Springs, CO, 1992).

[20] M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schelgel, K. Raghavachari, M. A. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Sceger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, S. Topiol, and J. A. Pople, Gaussian 90, Revision I (Gaussian, Inc., Pittsburgh, PA, 1990).

[21] M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schelgel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian 92, Revision A (Gaussian, Inc., Pittsburgh, PA, 1992).

[22] P. Flukiger, PhD Thesis, Universita of Geneva (1992).

[23] B. H. Besler, K. M. Merz, Jr., and P. Kollman, J. Comp. Chem. **11**, 431 (1990).

Distortion and Energetics in the Agonist Conformation of Bound Phenoxypropanolamine Agents in the β_1-Adrenoceptor

A. J. WILKINSON AND D. TIMMS

Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire,
England, SK10 4TG, United Kingdom

R. H. DAVIES AND K. J. BROADLEY

Welsh School of Pharmacy, University of Wales College of Cardiff, King Edward VII Avenue,
Cardiff, Wales CF1 3XF, United Kingdom

D. R. KELLY

Department of Chemistry, University of Wales College of Cardiff, King Edward VII Avenue,
Cardiff, United Kingdom CF1 3XF

Abstract

A theoretically contracted agonist conformation of potent phenoxypropanolamine derivatives on the β_1-adrenoceptor has been analyzed in detail. The main effect of the enthalpic contraction of some 6.0–7.0 kcal/mol arises from the movement of the nitrogen atom toward the aromatic ring by 0.7–0.8 Å, requiring some 3.0–3.5 kcal/mol. A second effect arising from the contraction can be a dihedral rotation of some 30° around the O—CH$_2$ bond of the planar anisole moiety. This rotation is correlated with an effect arising from "in-plane" deformation of the anisole moiety where opening of the relevant bond angle releases steric constraints for this rotation. *Ortho*-substituents assist this rotation indirectly through hyperconjugation with the lone pair of the OCH$_2$ group, electron-attracting substituents opening this bond angle and lowering the energy required to reach a given bond-angle deformation. The adjacent ring *meta*-substituent can be similarly affected, the strength of the total effect being also of the order of 3.0–3.5 kcal/mol. The net effect gives rise to a further contraction of the nitrogen atom and the beta-hydroxyl group toward the aromatic ring, the beta-hydroxyl group showing a contraction of up to 0.4–0.5 Å along the main axis of the conformer. The deformed conformation is consistent with the predicted conformer of a fixed-ring benzdioxepine molecule that possesses the highest degree of partial agonism within the set of phenoxypropanolamine agents. It is concluded that *ortho*-substituents in phenoxypropanolamine derivatives can retain steric freedom in both agonist and antagonist action provided that the substituent can accommodate the required deformation, both agonist and antagonist conformer forms lying within one unbound conformation. The agonist conformer is consistent with the proposed model for the ligand-activated transmembrane proton transfer in the β_1-adrenoceptor where a contraction along the main axis of the ligand conformer (with some attendant distortion in the position of the β-hydroxyl moiety) is required to activate proton transfer. © 1994 John Wiley & Sons, Inc.

1. Introduction

Comparative thermodynamics of the binding of agonists and antagonists in ligand–receptor complexes and in-site mutagenically modified receptors offer the

most accurate experimental evidence for gaining a detailed understanding of molecular mechanisms in signal transduction at the current time. Small perturbations to the binding ligand can throw light on the consistency of the binding mode of the ligand molecule with respect to the perturbations, on its conformation, and on the excitation of specific electrostatic and hydrogen-bond interactions, whereas in the region of the perturbing moieties, the effect of simple changes in the phase environment may be readily identified [1]. Under defined conditions, precise interatomic distance constraints (<0.5 Å) on receptor site atoms may be obtained if the conformation can be satisfactorily identified among other factors pertaining to the given mode of binding. These interatomic distance constraints are not absolute quantities but are relative to the mode of binding of the ligand and of the receptor.

In the case of the β_1-adrenoceptor, the identification of the phenolic oxygen atom of Tyr[377] as an electronegative atom capable of lying at a known geometric distance (10.3 ± 0.3) Å from an oxygen atom of Asp[138] in agonist action provided the basis for the development of an explicit transmembrane proton transfer model through α-helices III, IV, V, VI, and VII [2,3]. The interatomic distance constraint was based on a proposed agonist conformer of a phenoxypropanolamine molecule where a contraction along the main axis of the conformer [4] altered the form to approach the conformation of a known ethanolamine agonist isomer [5] (with some attendant distortion in the position of the β-hydroxyl moiety). The hypothesis of proton transfer through a Tyr[377]–Arg[156]–Tyr[157] proton shuttle led to a proposed alignment of the β_1-adrenoceptor α-helices on the bacteriorhodopsin model and the wider possibility that through a correlated gating mechanism an intermittent ion channel through helices I, II, III, and VII might be developed within the receptor [2].

Such a model should be consistent with available thermodynamic data on partially stimulating ligands bound to the β_1-adrenoceptor. Thermodynamic differences between agonist and antagonist binding modes may be explicitly calculated within the model, but a precisely determined form of the contracted phenoxypropanolamine agonist conformer is required for good accuracy. In this article, evidence is examined for the constraints on individual dihedral angles of the phenoxypropanolamine side chain, consistent with the overall energetics of contraction.

2. The β_1-Adrenoceptor Transmembrane Proton Transfer System

The CHARMM-based [6] model structures for the resting and activated states of the β_1-adrenoceptor using the agonist isoprenaline* were described previously. The structures are reproduced in Figure 1. In the resting state, Trp[183] restrains the Tyr[377] oxygen atom at the CHARMM distance of 5.2 Å from the Asn[373] amidic carbonyl group. A hydrogen-bond relay, Asn[369]–Ser[370]–Ser[145]–Asn[373], is held at its extremities by an Asn[369]–NH hydrogen-bond proton donor interaction with the carbonyl group of Pro[339] and by an Asn[373] carbonyl group interaction with Cys[336]. Asp[138] is the recognition site for the ligand amine moiety [7], the binding at this site, in turn, juxtaposing the β-hydroxyl group of the ligand with Asn[369]. Activation

---

* 2-(3,4-Dihydroxyphenyl)-1-isopropylaminoethan-2-ol.

of the Tyr^{377} phenolic oxygen atom is achieved by two hydrogen-bond proton donor interactions of the ligand, one directly by the *para*-hydroxy moiety of isoprenaline, the other indirectly by the beta-hydroxyl group's interaction with Asn^{369}, inducing rotation and hydrogen-bond reversal in the relay with rotation of Asn^{373}. Movement of the Tyr^{377} phenolic oxygen atom and shuttle system toward the establishment of the two hydrogen bonds with the rotated Asn^{373} and the *para*-hydroxyl group of the ligand liberates Trp^{183}, which moves to activate Tyr^{157}.

One interpretation for the proposed proton shuttle of Tyr^{377}–Arg^{156}–Tyr^{157} has the labile Tyr^{377} phenolic hydrogen atom interacting through a hydrogen-bond proton donor interaction with the unprotonated Arg^{156}, whereas a further proton donor interaction of the latter is to Tyr^{157}. An alternative is that Tyr^{377}–Arg^{156} exists as an ion pair near the cytoplasmic interface with Tyr^{157} protonated, the displacement and movement of Trp^{183} transferring the ion pairing to Arg^{156}–Tyr^{157}. Recovery can again utilize the position of Cys^{336}, which, in the activated state, can act as a catalyst for the recovery of the resting state through a transient role as the thienate ion [3].

The main features of the receptor site for the isoprenaline-activated state are summarized in Figure 2(a). A comparative CHARMM-optimized position for an uncontracted prenalterol conformer possessing an equivalent conformation to that for the isoprenaline is shown in Figure 2(b). Rotation of Asn^{373} produces a very weak NH–proton donor interaction with the ligand itself, and activation of Tyr^{377} is effectively weakened by one N—H · · · O Tyr^{377} interaction. Further movement of the proton shuttle is precluded without contraction of the phenoxypropanolamine molecule.

3. The Bound Agonist Conformation of β_1-Adrenoceptor Phenoxypropanolamine Partial Agonists—Energetics of "In-Plane" Distortion of the Phenoxy Moiety

Synthetic and pharmacological evidence has pointed to the side-chain conformation of ethanolamine and phenoxypropanolamine molecules being very similar in their bound agonist forms. The agonist conformation of flexible ethanolamine analogs is given by the fixed side-chain agonist isomer [4], whereas the known planarity of the anisole moiety in agonist and antagonist action [8], the lack of any electrostatic interaction of this group with the receptor, and the nonrelevance of folded intramolecular hydrogen-bonded forms in the agonist action of phenoxypropanolamine derivatives (phenoxypropanolamine and phenylaminopropanolamine analogs exhibit very comparable binding and partial agonism [9]—refer to Table I) reduce the likely active conformers to the equivalent ethanolamine form.

Contraction along the main axis of the phenoxypropanolamine molecule to initiate stimulus action in the β_1-adrenoceptor was based on the enthalpic differences observed between comparable phenethanolamine and phenoxypropanolamine agonist conformer interactions (5.5–7.0 kcal/mol) [4]. A theoretical 1.0 Å contraction in the position of the amine moiety toward the aromatic ring (with some distortion in the position of the β-hydroxyl moiety) retaining the same conformation required an estimated minimal basis STO-3G energy of 4.6 kcal/mol. Differences in vibra-

WILKINSON ET AL.
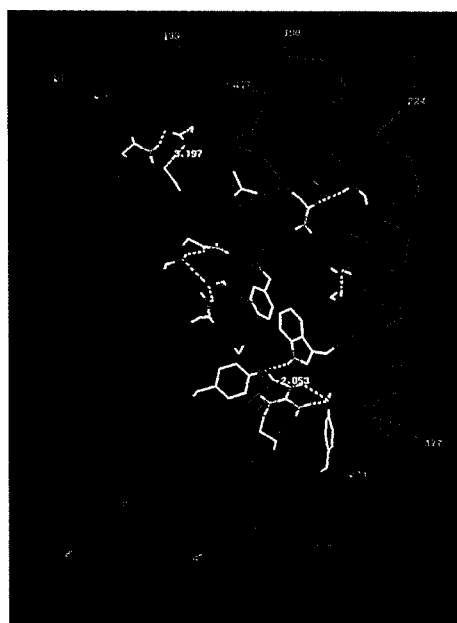
(a)
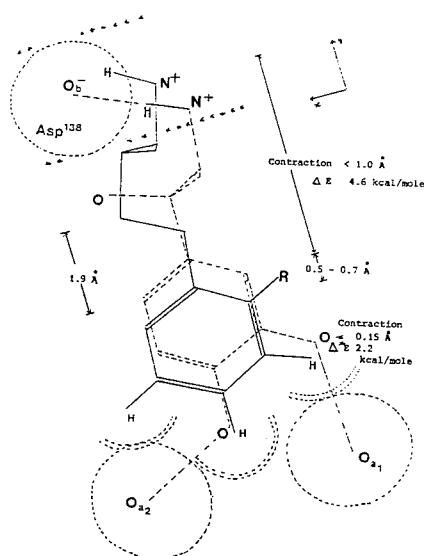
(b)
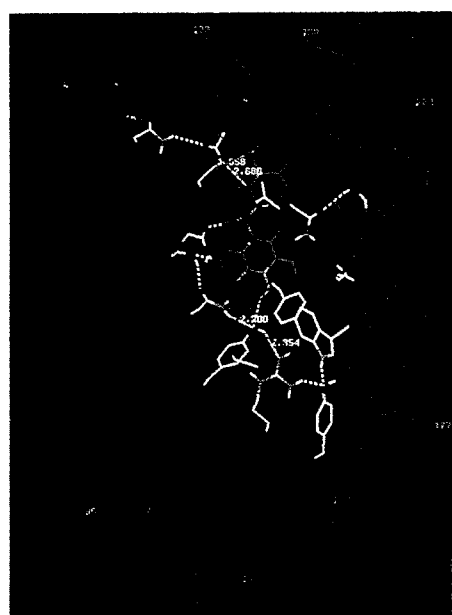
(c)

(d)

tional contributions to the binding between the relaxed and contracted forms were negligible.

These calculations were performed on the vinyl ether analog conformer using the unprotonated molecule in view of the charge compensation from the ion-pair interaction at Asp^{138}. The main effect arises from rotation about the bond governing the position of the nitrogen atom, bringing this atom to an eclipsed position with the hydrogen atom H_1 [Fig. 3(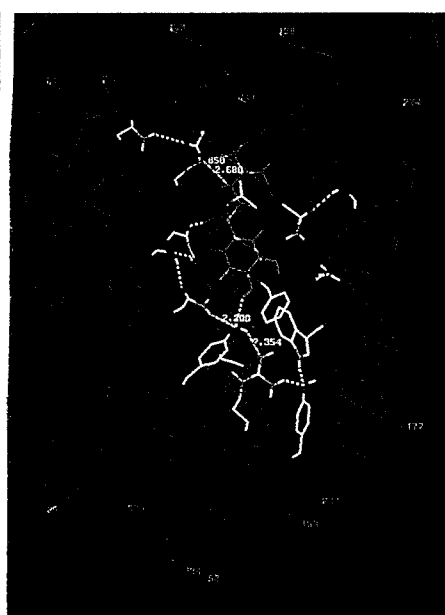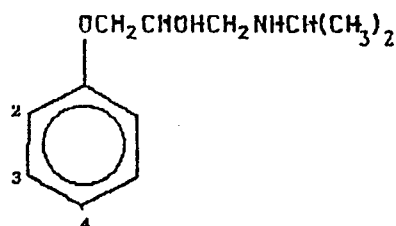a)]. The 1.0 Å contraction is slightly too large, the nitrogen atom coming past the eclipsed form into part of the conformer space of a nonrelevant intramolecular hydrogen-bonded form, but in the eclipsed form, an overall contraction of 0.7–0.8 Å would require 3.0–3.5 kcal/mol. If the ethanolamine agonist conformation is strictly maintained by constraining dihedral angles, a contraction of only 0.3 Å is achievable for an STO-3G expenditure of energy of 5.5 kcal, the three bond angles governing this contraction showing a general narrowing by 4.0–4.5°.

There remains a 2.5–3.5 kcal/mol enthalpy to be accounted for in the energetics of the conformer contraction. Within the model, and the known planarity of the anisole moiety, the contraction can be associated with an expected "in-plane" deformation of the aromatic phenoxy moiety. Two effects may be anticipated from this distortion: One is model-dependent on the siting of the second electronegative receptor site atom, giving rise to the approximately one-dimensional conformer contraction [Fig. 1(b)]. The positioning of this atom was based on the known interactions of practolol [10], prenalterol, and pindolol [4] and on the van der Waals contacts of *meta-* and *para-*hydrogen atoms of *ortho-*substituted partial agonists. (Molecular structures are identified in Table I. Practolol is the $4-NHCOCH_3$ analog). Moderate "in-plane" distortion of the *meta-* and *para-*hydrogen atoms was anticipated from the contraction. The second effect arises from "in-plane" deformation of the planar—OCH_2 group itself, which can affect the orientation of the aromatic moiety with respect to the propanolamine side chain. More importantly, for achieving a required conformation with a planar—OCH_2 moiety, it facilitates moderate movement of rotation about the C_7—O_3 bond [Fig. 3(b)] by opening the relevant bond angle in the undeformed molecule with the consequent slight lowering in energy required to reach the given conformation. Since both effects are sensitive to the presence of an *ortho-*substituent, the influence of this group on the aromatic

Figure 1.   (a) Resting state of the β_1-adrenoceptor. Side view showing the residues involved in the proposed mechanism of proton transfer. Trp^{183} holds Tyr^{377} at an oxygen–oxygen atom distance of 5.2 Å from Asn^{377}. The three residues near the cytoplasm are Tyr^{377}, Arg^{156}, and Tyr^{157}. (b) Superimposition of phenethanolamine (dotted) and phenoxypropanolamine conformers within the β_1-adrenoceptor site. The estimated enthalpy of contraction of the phenoxypropanolamine agonist conformer to interact with two electronegative receptor sites O_a and O_b is some 6–7 kcal [4]. Two possible positions (1 and 2) exist for the receptor oxygen atom O_a where the atom is predicted to be in van der Waals contact with the *meta-* and *para-*hydrogen atoms of the *para-*unsubstituted phenoxypropanolamine ligand (the β-hydroxyl group of the latter has been omitted for clarity). (c,d) Isoprenaline–β_1-adrenoceptor activated state showing reversal of the $Asn^{369}-Ser^{370}-Ser^{145}-$ Asn^{373} hydrogen-bond relay and movement of Trp^{183} to hydrogen bond with Tyr^{157}.

(a)   (b)

Figure 2. (a) Positions of the hydrogen-bond relay Asn^{367}–Ser^{370}–Ser^{145}–Asn^{373} and of
the proton shuttle Tyr^{377}–Arg^{156}–Tyr^{157} in the isoprenaline conformer β_1-adrenoceptor
activated state. (b) Relative positions of the hydrogen-bond relay and proton shuttle for
the isoprenaline conformer activated state (white) and for an equivalent prenalterol con-
former in the extended form (yellow) in one possible antagonist mode of binding.

ring deformation is examined in detail. The *ortho*-substituents in sets of pheny-
oxypropanolamine agents provide ready control of the degree of stimulant activity
within β_1-adrenoceptor complexes, and this influence has been the subject of earlier
conformational hypotheses [1,12].

Two-substituted phenols were considered an adequate model for assessing the
salient features of "in-plane" aromatic ring distortion. The primary effect of dis-
tortion of aryl hydrogen atoms in the x–y plane of the aromatic ring of phenols
and anisoles is that of σ-bond distortion. For small distortion, bond lengths will
change little, and the π-bond overlap will be relatively unchanged. The effects of σ
distortion should be reproduced, therefore, even with limited basis calculations.
Effects of overestimation of π donation with a vacant p orbital effectively expanding
the basis set available for π-electron stabilization [13] will have no direct relevance
to the deformation, but could have secondary effects through changes in intrinsic
bond lengths. STO-3G calculations on experimental mesomeric effects are in good

TABLE I. Stimulus response and binding of some phenyoxypropanolamine derivatives on the cardiac β_1-adrenoceptor:

$$OCH_2CHOHCH_2NHCH(CH_3)_2$$

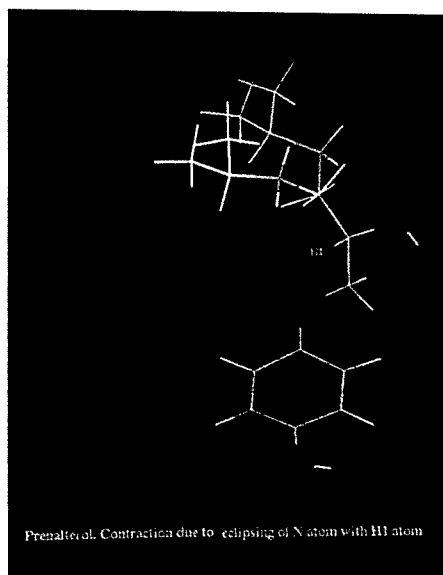| Compound		$\delta\Delta G_L$[a] (kcal)	i.s.a.[b] (heart beats/min)	Stimulus function $-RT \log [e/(l-e)]$[c] (kcal)
4-OH	(Prenalterol)	−0.8	129 ± 5	
	Benzdioxepine	Not known	151	
4-CH₃		+2.7	Nonstimulating	
2-F		0	117 ± 2	−0.02 ± 0.04
2-H		0	104 ± 7	+0.12 ± 0.075
2-OCH₃		0	101 ± 7	+0.15 ± 0.075
2-CF₃		0	90 ± 5	+0.27 ± 0.06
2-NO₂		0	66 ± 5	+0.56 ± 0.065
2-CH₃		0	65 ± 11	+0.575 ± 0.15
2-C₂H₅		0	29 ± 7	+1.19 ± 0.17
2-COCH₃		0	74 ± 5	
2-COCH₃ ⎫	Phenylamino-	∼0	61	
2-CONH₂ ⎭	propanolamine[a]	∼0	76	

[a] The relative free energies of binding $(-RT \log K_L)$ are referenced to a nonaqueous environment using the long-chain ester PGDP/water model [11].

[b] Intrinsic sympathomimetic activities (i.s.a.) and the resultant stimulus functions are given for the rat heart. The maximum incremental response (e_A) is ∼230 beats/min [8].
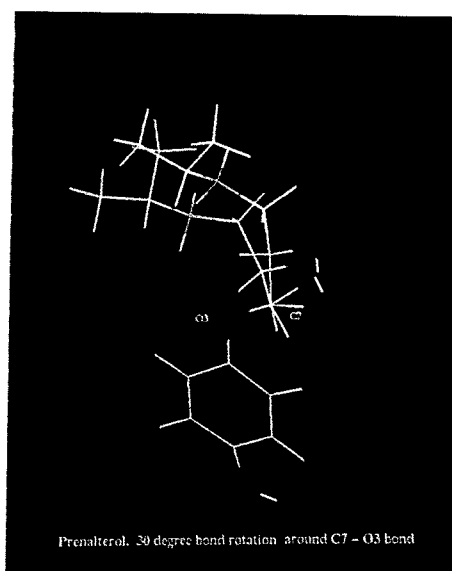
[c] $e = e_B/e_A$, where e_B and e_A are the maximum stimulatory responses of the partial and full agonists, respectively.

agreement in monosubstituted benzenes [14], whereas methoxy group nonplanarity in *ortho*-dimethoxybenzenes is also well represented in this basis [15].
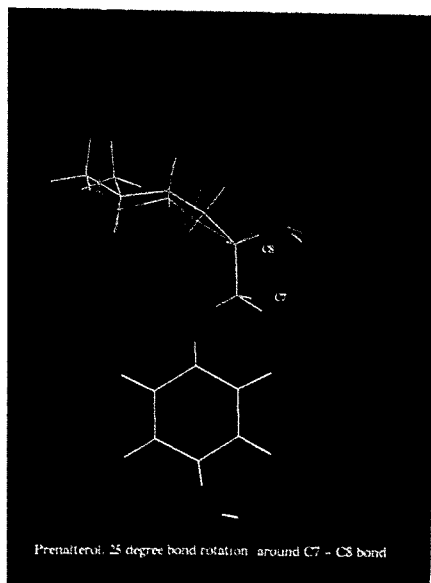
All molecular structures, subject to the given constraints, were fully optimized at the HF–SCF level using the Gaussian suite of programs. Early work with 4-31G and STO-3G basis sets employed Gaussian 80 and Murtagh–Sargent optimization. Structures and harmonic force field data for the 2-CH₃, 2-H, and 2-F analogs together with calculations at the 6-31G** level and data for substituted anisoles utilized analytic gradient techniques with Gaussian 82 and 88 on a Convex C220 machine [16]. Bond lengths and bond angles were determined to < 0.0005 and < 0.1, respectively. In all calculations on phenol derivatives, the planarity of the phenol skeleton was retained with the phenolic hydrogen atom fixed at a bond angle of 115°, similar to the value found in phenoxymethyl moieties [15].
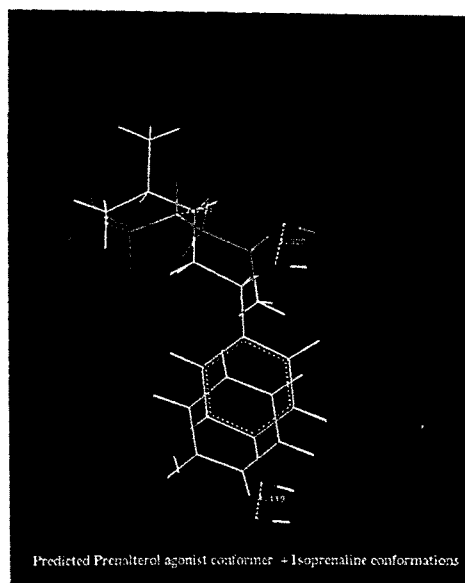
WILKINSON ET AL.

Prenalterol. Contraction due to eclipsing of N atom with H1 atom

Prenalterol. 30 degree bond rotation around C7 – O3 bond

(a)   (b)

Prenalterol. 25 degree bond rotation around C7 – C8 bond

Predicted Prenalterol agonist conformer + Isoprenaline conformations

(c)   (d)

Substituted phenolic structures were optimized subject to the following constraints: (1) none, (2) 3-H atom distorted 15° (the labeling of hydrogen atoms given from the 2-substituent), (3) the 3-H and 4-H atoms both distorted 7°, (4) the 5-H atom distorted 15°, (5) the 4-H and 5-H atoms both distorted 7°, and (6) the 3-H and 5-H atoms both distorted 7° (Tables II and III). Table IV gives optimized geometries for 2-substituted anisoles and for marginal "in-plane" deformation of the 1-OCH$_3$ group.

The energetics of distortion at the free-energy level contains enthalpic and entropic contributions from vibrational differences between the deformed and undeformed structures, together with a small difference in the zero-point energies. For the 2-H and 2-CH$_3$ phenol conformers, relevant differences in the STO-3G normal vibrational energy modes of the two forms were found to be small ($\Delta \sim 10$ cm^{-1}) and, as expected, the lowest vibrational normal modes (260 cm^{-1}) were too high for a significant entropic difference between the deformed and undeformed vibrational forms. Vibrational energy differences have therefore been neglected. An analysis of the vibrational modes in the phenoxypropanolamine side chain [4] showed that in spite of entropic contributions from much weaker vibrational modes the STO-3G estimate of the difference in the vibrational entropy between the deformed and undeformed species was only 0.15 kcal/mol.

The deformational energies are very comparable with all three basis sets. In the STO-3G minimal basis set with the 2-NO$_2$ group, energy optimizations gave very shallow minima. In the unconstrained molecule, the —NO$_2$ group was orientated 21° to the plane of the aromatic ring, whereas in the constrained form (2), this orientation was reduced to 5° and effective planarity in agreement with the unconstrained 2-nitro anisole structure (Table IV). Comparable X-ray data suggest larger variation (52° [17,18] and 32° [19]) in the effective orientation of the nitro group under the influence of crystal forces. General trends here may not be well represented.

Table II shows that electronic influences of the *ortho*-substituent on the deformational energy of 3H, and 3H, 4H in-plane aromatic ring distortion to be dominant. For the three conformers of 2-CH$_3$ phenol, the similarity in the STO-3G deformational energy of 6.3 kcal for the constrained form (2) indicates that steric contact between the substituent and the phenolic oxygen atom is not an intrinsic problem in these forms, in agreement with van der Waals contact distances. Marginal steric contact in the deformed 2-CF$_3$ dominant conformer may have an influence

Figure 3. Contracted agonist conformer in phenoxypropanolamine binding. (a) The predominant effect of a 3.0–3.5 kcal/mol enthalpic contraction (yellow) on the prenalterol conformer equivalent to that for isoprenaline. In the contracted conformer, the nitrogen atom is eclipsed to hydrogen atom H$_1$. (b) Indirect effect (pink) of a 3.0–3.5 kcal/mol enthalpic "in-plane" deformation of the planar—OCH$_2$ moiety on the eclipsed conformer (a) showing a 30° rotation about the C$_7$—O$_3$ bond. (c) The result of a 26° dihedral increment around the C$_7$—C$_8$ bond (green) to minimize differences from the isoprenaline staggered conformer. (d) Comparison of relevant intermolecular distances in molecular overlap between the expected agonist conformers for prenalterol and isoprenaline.

WILKINSON ET AL.

TABLE II. Geometry-optimized 4-31G and STO-3G structures for 2-substituted phenols with given "in-plane" deformational constraints on 3- and 4-substituents [(1) v, w, optimized; (2) v = 135°; (3) v = 127°, w = 127°]:
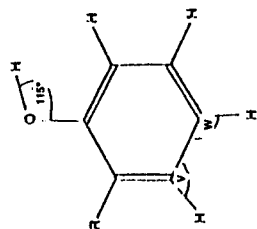
R	4-31G (au)			STO-3G (au)				
	(1)	(2)	$\Delta E_{2\text{-}1}$ (kcal/mol)	(1)	(2)	(3)	$\Delta E_{2\text{-}1}$ (kcal/mol)	$\Delta E_{3\text{-}1}$ (kcal/mol)
H	-305.12501	-305.11632	-5.45	-301.72830	-301.71879	-301.72419	+6.0	+2.5
F	-403.84870	-403.84220	+4.1	-399.18374	-399.17589	-399.18005	+4.9	+2.3

−340.31267	−340.30254	−340.30816	+6.4	+2.8
−340.312105	−340.302055	−340.30770	+6.3	+2.8
−340.31232	−340.30234	−340.30798	+6.3	+2.7
−632.69566	−632.68708	−632.69190	+5.4	+2.4
−632.69276	−632.68327	—	+5.9	—
−414.14346	−414.13276	—	+6.7	—
−414.14310	−414.13483	−414.13935	+5.2	+2.35
−502.43115[b]	−502.42466[c]	−502.42822	+4.1	+1.8

[a] For $v = 130°$, $\Delta E_{2-1} = 3.0$, 2.7, and 2.1 kcal for R = CH_3, H, and F, respectively.

[b] NO_2 group orientation 21° to ring plane with shallow minimum.

[c] NO_2 group 5° to plane of ring but equivalent energy found with planarity constraint (energy difference < 0.01 kcal).

TABLE III. Geometry-optimized 6-31G** and STO-3G structures for 2-substituted phenols with given "in-plane" deformational constraints on 4- and 5-substituents [(1) $v'w'$ optimized; (4) $v' = 135°$; (5) $v' = 127°$, $w' = 127°$; (6) $v' = 127°$, $w = 127°$]:
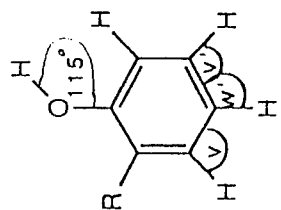
| R | 6-31G** (au) | | $\Delta E_{5\text{-}1}$ (kcal/mol) | STO-3G (au) | | | $\Delta E_{4\text{-}1}$ (kcal/mol) | $\Delta E_{5\text{-}1}$ (kcal/mol) |
	(1)	(5)		(1)	(4)	(5)		
H	−305.57291	−305.56934	2.2	−301.72830	−301.71901	−301.72440	5.8	2.4
F	−404.41599	−404.41281	2.0	−399.18374	−399.17465	−399.18018	5.7	2.2

|						$\Delta E_{6\text{-}1}$	
−344.61331	−344.61008	2.0	−340.31267	−340.30374	−340.30905	5.6	2.3
−641.19113	−641.18771	2.15	−632.69566	−632.68634	−632.69179	5.85	2.4
				(1)	(6)	2.8	
				−301.72830	−301.72381	2.3	
				−399.18374	−399.18006		
			−340.31267	−340.30809		2.9	
			−632.69566	−632.69156		2.6	

TABLE IV. Geometry-optimized STO-3G structures for 2-substituted anisoles with given "in-plane" deformational constraints on the 1-OCH$_3$ substituent [dihedral angle τ 14-8-7-6; (7) x optimized; (8) x = 113°; (9) x = 109°]:
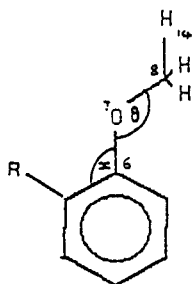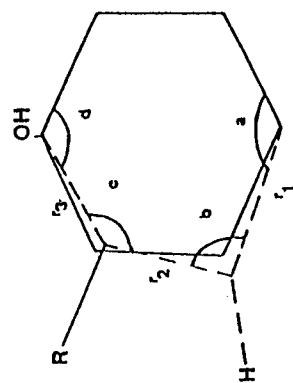

R	(au) (7)	$x°$	$\theta°$	(au) (8)	$x°$	$\theta°$	ΔE_{8-7} (kcal/mol)
H	−340.31184	114.6	114.8	−340.31166	113	114.6	0.1
F	−437.76782	114.3	114.4	−437.767705	113	114.3	0.1
H H—C—H H	−378.89605	114.2	115.0	−378.89593	113	114.8	0.1
NO$_2$	−541.01434	117.3	115.1	−541.01296	113	114.5	0.9

H	(au) (9)	$x°$	$\tau°$				ΔE_{9-7} (kcal/mol)
(a)	−340.30957	109	180.0 (optimized)				1.4
(b)	−340.30667	109	150.0				3.2
(c)	−340.30465	109	135.0				4.5
(d)	−340.30390	109	120.0				5.0

on the energetics between the deformed and undeformed species, but the results show that this influence is small and of the order of 0.5 kcal despite the exaggerated bond lengths predicted by the minimal basis. The comparative difference between the 2-F and 2-CH$_3$ deformational energies is 1.4 kcal, indicating that electronic influences dictate the deformational energy of the 2-CF$_3$ compound. The correlated but weaker trend is repeated in the constrained form (3).

Table III shows the influence of the 2-substituents on 4-H and 5-H in-plane bond-angle distortion to be negligible, whereas differences in Tables II and III again indicate a general trend in the ease of ring deformation with decreased electron ring density for the 3-H and 3-H, 4-H deformations. The underlying causes of this trend may be examined. The primary bond angle and bond length changes in the aromatic ring of the constrained form (2) are given in Table V. Bond-distance changes at r_1 and r_3 for these small perturbations are very small, being ~0.003 and −0.005, respectively, for all molecules. Leading bond-angle changes within the aromatic ring are as expected and are very similar for both extended split-valence and minimal basis sets. The primary internal bond-angle narrowing produced in

TABLE V. Calculated bond-angle and bond-length distortions in the aromatic ring for structure (2) with $v = 135°$ (u, undeformed; d, deformed):

R		$a°$	$\Delta a°$	$b°$	$\Delta b°$	$c°$	$\Delta c°$	$d°$	$\Delta d°$	r_2 (Å)	Δr_2 (Å)
4-31G results											
F	u	119.5		119.5		121.8		118.5		1.369	
	d	120.7	+1.3	116.7	−2.8	123.9	+2.1	118.0	−0.5	1.382	+0.013
H	u	119.4		120.6		119.5		120.5		1.380	
	d	120.9	+1.5	117.3	−3.3	123.6	+4.1	119.9	−0.6	1.396	+0.016
STO-3G results											
F	u	120.0		120.2		120.0		119.1		1.387	
	d	121.4	+1.4	117.1	−3.1	122.3	+2.3	118.6	−0.5	1.399	+0.012
OCH₃ᵃ	u	119.8		121.0		119.0		119.6		1.387	
	d	121.2	+1.4	117.8	−3.2	121.5	+2.5	119.1	−0.5	1.401	+0.014
CF₃	u	119.5		120.6		119.7		119.6		1.384	
	d	121.1	+1.6	117.3	−3.3	122.2	+2.5	119.2	−0.4	1.398	+0.015
H	u	119.6		120.6		119.7		119.8		1.381	
	d	121.2	+1.6	117.1	−3.5	122.3	+2.6	119.3	−0.5	1.397	+0.016
CH₃ᵃ	u	119.7		121.5		118.2		120.6		1.384	
	d	121.5	+1.8	117.8	−3.7	120.8	+2.6	120.2	−0.4	1.401	+0.017
NO₂	u	119.3		120.1		121.3		117.8		1.388	
	d	120.5	+1.2	117.5	−2.6	123.2	+1.9	117.3	−0.5	1.401	+0.013

ᵃ Preferred deformed conformer. Refer to Table II.

the *meta* position (−Δ*b*) has compensation with expanded adjacent angles (a) and (c) with a further bond-angle (d) contraction. Although there is a trend in increased deformation with decreased σ-electron density, the intrinsic effects of substituents on bond angles are more marked.

Figures 4 and 5 show the relative orientation of the ring substituents for the deformed and undeformed species. A hyperconjugative interaction in the σ-bond framework is evident. Figure 4 shows the tilts of the phenolic oxygen atoms given
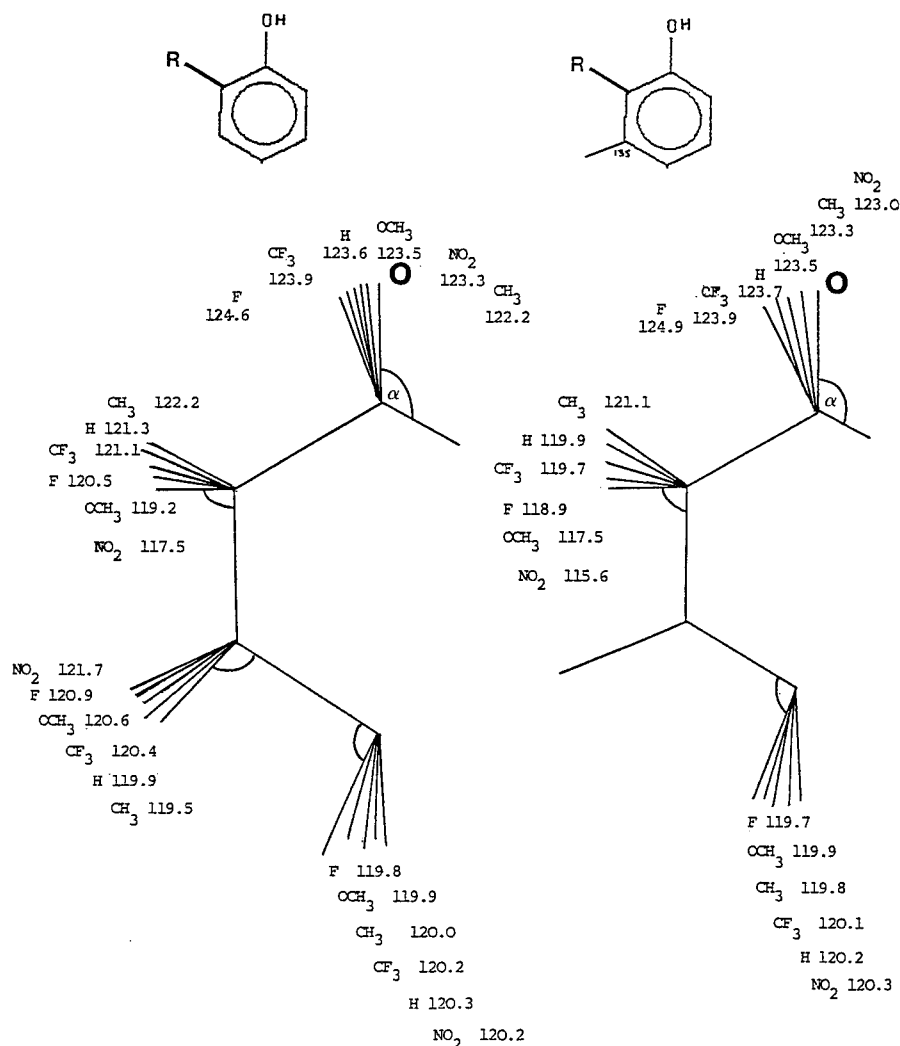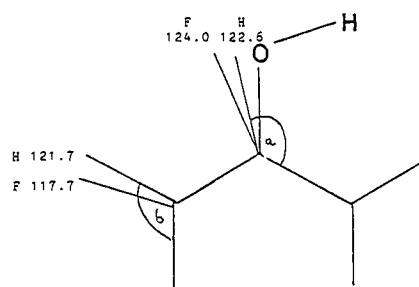

Figure 4.   STO-3G estimates of bond angles for aromatic ring substituents in 2-substituted phenols comparing conformer (2) where the relevant *meta*-hydrogen atom is constrained at 135°, with the fully optimized phenol conformer.

4-31G

F 124.0   H 122.6   H
O
α

H 121.7
F 117.7
b

STO-3G

O — H
u 1.381
d 1.382

u 116.62
d 116.87
u 123.64
d 123.75

H
u 118.98
d 117.84

u 119.74
d 119.38

u 121.31
d 119.91
u 119.71
d 122.25

u 1.397
d 1.392

O — H
u 1.382
d 1.382

u 116.07
d 116.57
u 124.63
d 124.88

F
u 119.68
d 118.75

u 119.30
d 118.55

u 120.54
d 118.88
u 119.78
d 122.37

u 1.409
d 1.404

O — H
u 1.382
d 1.383

u 117.17
d 116.46
u 122.23
d 123.31

CH₃
u 119.58
d 118.16

u 120.60
d 120.33

u 122.22
d 121.11
u 118.20
d 120.73

u 1.405
d 1.399

O — H
u 1.378
d 1.378

u 116.75
d 116.94
u 123.91
d 124.05

CF₃
u 118.83
d 117.92

u 119.34
d 119.01

u 121.16
d 119.71
u 120.01
d 122.37

u 1.400
d 1.395

O — H
u 1.383
d 1.383

u 116.86
d 117.20
u 123.52
d 123.68

OCH₃
u 121.73
d 121.03

u 119.62
d 119.12

u 119.23
d 117.51
u 119.04
d 121.46

u 1.410
d 1.404

O — H
u 1.371
d 1.372

u 118.79
d 119.59
u 123.34
d 123.08

NO₂
u 121.26
d 121.16

u 117.87
d 117.33

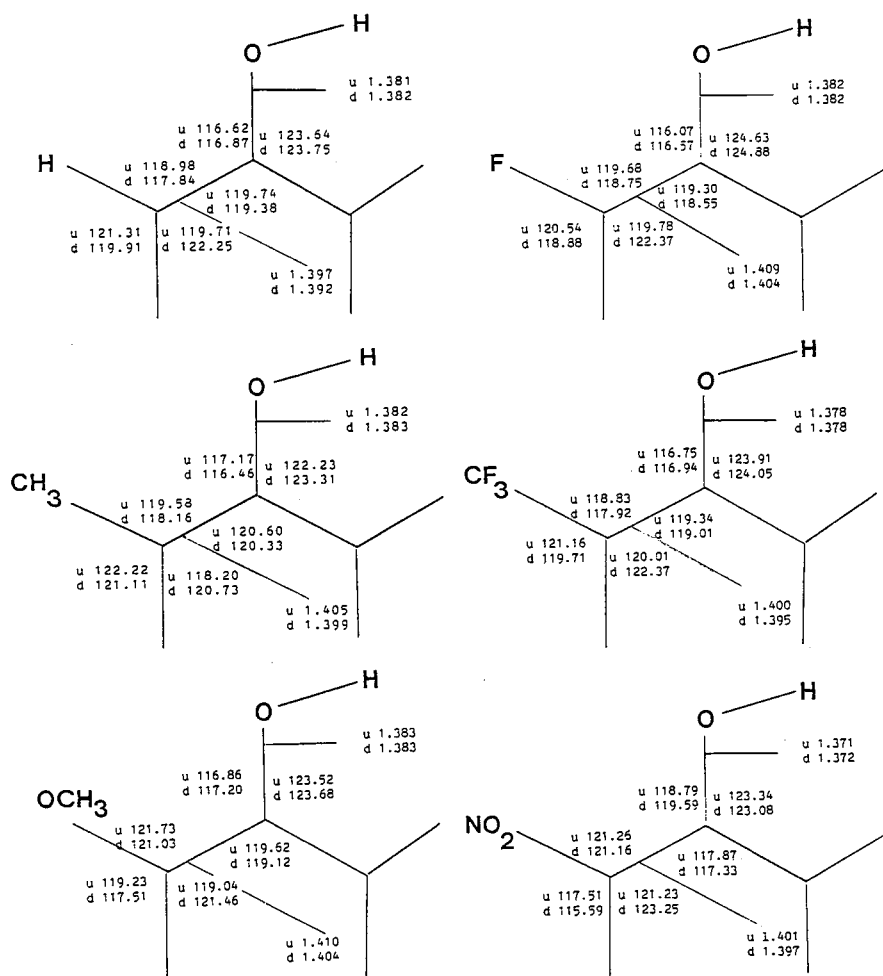u 117.51
d 115.59
u 121.23
d 123.25

u 1.401
d 1.397

Figure 5. 4-31G estimates for bond angles in the constrained conformer (2) of 2-substituted phenols, and selected STO-3G bond lengths and bond angles for the constrained (d) and fully optimized (u) conformers.

by angle α from the simple sp^2 hybridization value, where reduced electron density in the σ-bond framework increases the angle α due to the lone-pair overlap of the oxygen atom with the adjacent vacant antibonding C—C* orbital [20]. The ranking of the CH_3, H, CF_3, and F derivatives are consistent with this effect and the relative orientation of the 2-substituent itself is similarly affected. The 4-31G data on the 2-H and 2-F structures give very similar results. The results indicate that weak electrostatic repulsions between the 2-substituent group and the phenolic oxygen atom do not affect the ranking. The positions of the 2-OCH_3 and 2-NO_2 derivative are more difficult to assess. Both 2-OCH_3 and 4-OCH_3 anisoles have facile rotation of the methyl groups out of the plane of the ring, and the nonplanar conformers have similar populations to the planar forms [15], but close steric contact of the *ortho*-nitro group with the methoxy moiety is shown by marginal "in-plane" movement of the anisole moiety toward the *ortho*-substituent (Table IV). Taking the mean "in-plane" tilt of the undeformed aromatic ring (i.e., the difference in the external bond angles affecting the oxygen atom) as a guide to the overall effect gives the tilt angles for F of 4.1°; H, 3.5°, OCH_3, 3.3°; CF_3, 3.6°; NO_2, 2.5°; and CH_3, 2.5°.

The effect of deformation affecting the 4-H atom is weak, but consistent with the expected trend. In the constrained conformer (3) in Table II, expansion of the *para*-hydrogen atom bond angle w, with the consequent narrowing of the bond angle a ($-\Delta a$), would create a compensatory effect on b ($+\Delta b$). Thus, electronegative 2-substituents would have a reduced effect due to this compensation. Within the limits of the basis, some evidence for this is shown by the position of the 2-H derivative in the relative energetics of the 2-CH_3, 2-H, and 2-F constrained forms. The 2-H derivative energy is relatively enhanced. This effect is confirmed by deformation of the *para*-hydrogen atom alone. For $w = 130°$, the deformational energies are for 2-F, 2.7_5; 2-CH_3, 2.6_4; CF_3, 2.5_3; and H, 2.5_2 kcal. Although the influence of the 2-substituent is now very marginal, it is again striking that the bond angle of the *para*-hydrogen atom in the fully optimized molecule correlates with the deformational energies. Deformed structures of conformer (2) with the same bond-angle constraint ($v = 130°$) show no relative change in the pattern of *meta*-hydrogen atom deformation.

With these deformational patterns, it is seen that the effects of "in-plane" deformational movements of the 1- and 3-substituents toward the 2-*ortho*-substituent will correlate for small positional changes, but in the case of the 2-nitro group, repulsive proximity effects are apparent even in the optimized form. Table IV shows the STO-3G energetics for rotation about the planar O—C bond of the 1-OCH_3 moiety in anisole, for some 6° "in-plane" deformation of the group toward the 2-H substituent. The "in-plane" deformation requires some 1.4 kcal/mol, whereas a rotation of the dihedral angle up to 30° requires a further enthalpy of 1.8 kcal/mol, giving a total of 3.2 kcal/mol.

4. The Bound Agonist Conformation of β_1-Adrenoceptor Phenyoxypropanolamine Partial Agonists—Side-chain Conformation

Data "*in vitro*" on the stimulant response of a number of phenoxypropanolamine derivatives are given in Table I. Such data can be complicated by the signal am-

plification of the responses [21], but evidence that this factor can remain constant within given sets of β_1-adrenoceptor binding data will be given in a separate article [22]. The stimulatory response function in Table I may be written

$$-RT \log \frac{e_B/e_A}{1 - e_B/e_A} = -RT \log \frac{K^{ij^x}}{K^{rs^x}} -RT \log (\tau + 1) , \qquad (1)$$

where the maximum stimulatory response e_B of the partial agonist B relative to that of a full agonist A is related to the ratio of the agonist conformer (a) and antagonist (r) receptor binding constants, the j^x and s^x indicating the appropriate receptor conformation in the binding. τ is the amplification factor. For high signal amplification, the observed binding constant is dominated by the antagonist component, and for the *ortho*-substituent set under examination, the binding constants are invariant when referenced to an effective hydrocarbon environment [10] (± 0.12 kcal/mol). The agonist response on the free-energy scale indicates a very sensitive effect, one that gives an increment of 10 heart beats/min on the rat heart rate for a free-energy change of little more than 0.1 kcal/mol. The comparative model for the agonist activity of 2-substituted phenoxypropanolamine derivatives is, however, one of particular simplicity. The compression of the alphatic side chain is effectively a constant of the comparison except for the effects produced by deformation of the anisole moiety. Electrostatic interactions of the aromatic ring hydrogen atoms with the adjacent receptor hydrogen-bond proton acceptor group are not expected to be detectable (even the strongest of $O \cdot \cdot \cdot CH_3$—X interaction, where X is halogen or equivalent [23], here, would not contribute more than 20% of the overall variation in the stimulus observed). The 2-substituted conformers at the point of initiating the stimulus thus have effectively identical interactions to the two electronegative sites, whereas the antagonist contributions are also invariant. Under such conditions, the observed biological variation is expected to correlate with the fraction of deformed conformer but coupled to constant signal amplification. Since the 2-substituents in the bound agonist form exist in a local nonpolar hydrocarbon environment, variations in the deformed conformer populations are almost certainly due to intrinsic properties of the molecular species.

Figure 6 shows the hyperbolic response function of the *ortho*-substituted phenoxypropanolamine derivatives plotted against the relative agonist to antagonist conformer populations for *meta*- and *para*-hydrogen atom distortion on the thermodynamic scale at 310°K. The relative conformer populations have been based on the intrinsic STO-3G energies of the deformed and unconstrained species using the phenol model. There is an obvious structural constraint on the 2-C_2H_5 agonist conformer and its population has been taken as that of the 2-CH_3 form adjusted for a restriction in the equivalent overlapping *meta* region of the isoprenaline conformer, giving an entropic change of 0.48 $\log_{10}$ units.

This assumption is in agreement with an energetics of deformation of the 3-*meta*-hydrogen atom toward the 2-substituent in the region of 2.3 kcal/mol. A similar smaller symmetric "in-plane" deformational effect of the 1-OCH_3 group of some 1.5 kcal/mol will also correlate with this effect. Although it is not possible to distinguish either effect separately, they account for the unusually high agonist
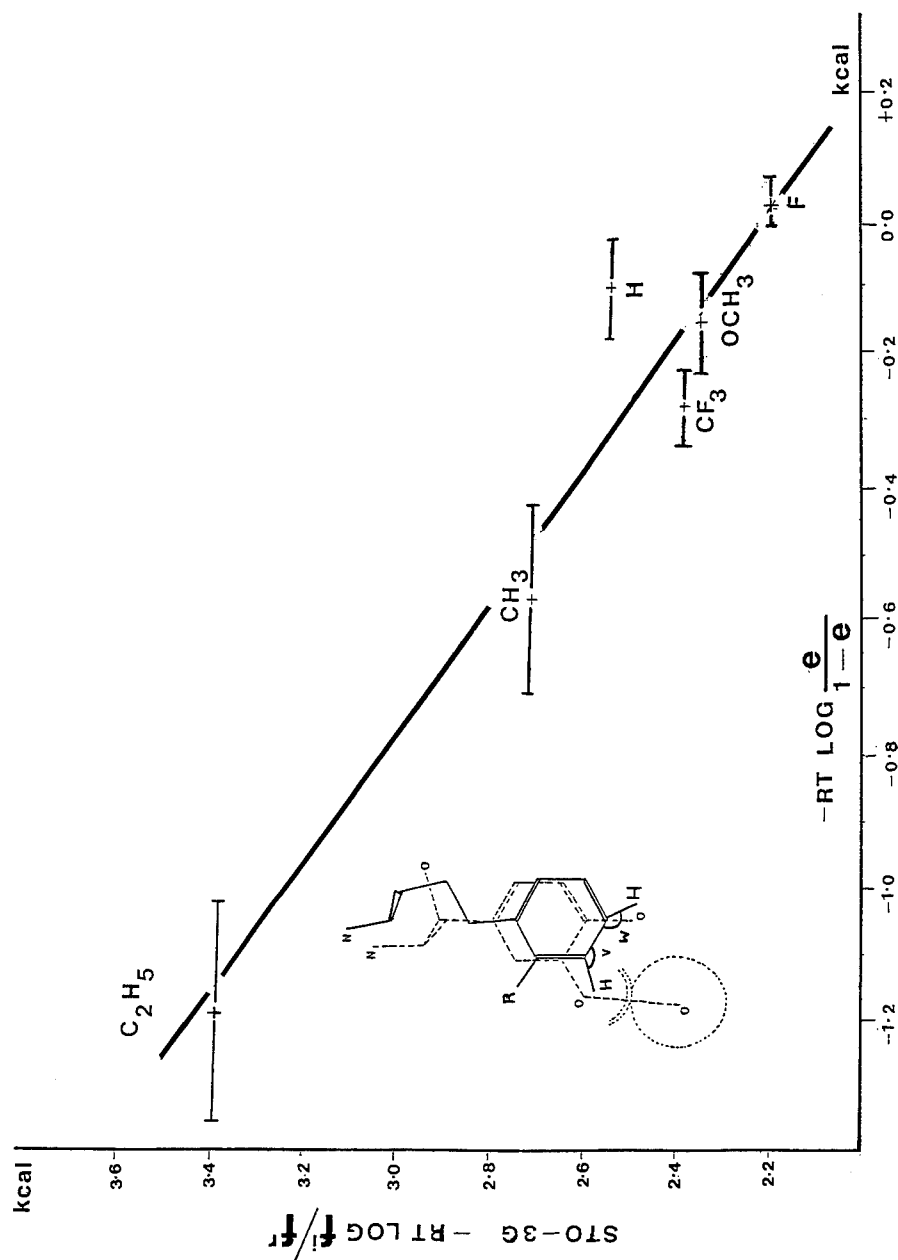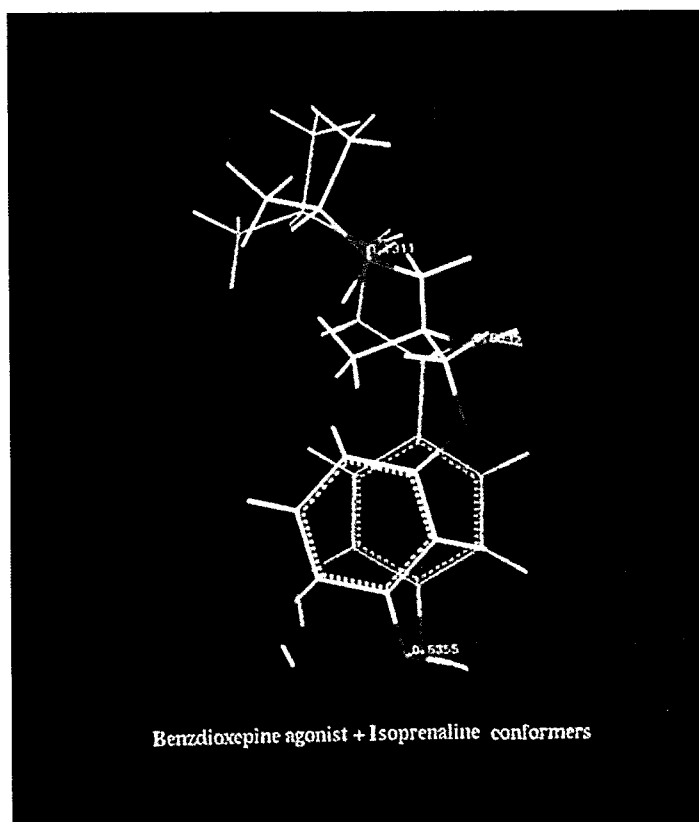
Figure 7.   Comparison of intermolecular distances in molecular overlap between the predicted agonist conformers of the benzdioxepine molecule and isoprenaline.

action shown by the relatively large $2\text{-}CF_3$ substituent coupled with the general observation that small *ortho* groups are generally associated with higher intrinsic stimulant activity in this set of compounds. The failure of the planar 2-nitro group to exhibit maximum stimulation in the corresponding analog argues for the dominant effect of "in-plane" deformation of the $1\text{-}OCH_3$ group in the bound agonist

Figure 6.   The hyperbolic response function for 1-isopropylamino, 3-(2R) phenoxy propan-2-ol derivatives on the rat cardiac β_1-adrenoceptor plotted on the thermodynamic scale against the free-energy component for moderate distortion of the aryl *meta-* and *para-* hydrogen atoms some $7°$ from their standard positions. The STO-3G conformer population of the distorted conformer i with bond angles $v = 127°$ and $w = 27°$ relative to the undistorted form r is modeled on 2-substituted phenol conformers. e is the intrinsic stimulant activity of the compound. The free-energy component for $R = C_2H_5$ has been taken as that for $R = CH_3$ with an additional unfavorable entropic $T\Delta S$ contribution of 0.68 kcal/mol at $310.15°K$.
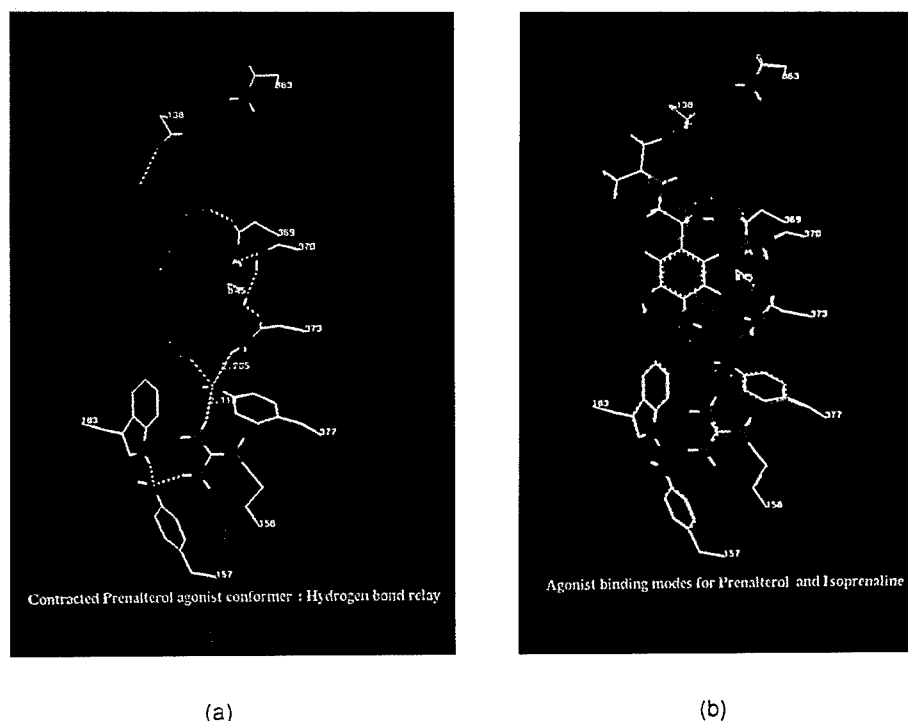
(a)   (b)

Figure 8.   (a) Hydrogen-bond distances in the activated form of the proton transfer system
for the contracted prenalterol conformer in the ligand complex. (b) Relative positions of
the hydrogen-bond relay and proton shuttle for the activated receptor complex with the
contracted prenalterol conformer (yellow) and with the isoprenaline agonist conformer.

state. A guide to the net effect of this deformation can be given from the STO-3G
estimates of the mean "in-plane" tilt angle of the undeformed aromatic ring.

5. The Bound Agonist Conformation of β_1-Adrenoceptor Phenoxypropanolamine Partial Agonists—Consistency with the Transmembrane Proton Transfer Model

Figure 3(b) shows the effect of a rotation of 30° around the C_7—O_3 bond for an
enthalpic "in-plane" deformation of some 6° of the 1-OCH$_3$ group toward the 2-
substituent in the case of prenalterol. The STO-3G estimate of the enthalpic incre-
ment for the contraction given in (a) and (b) is some 6.0–7.0 kcal/mol. To attain
the closest structure to the staggered ethanolamine agonist conformation, the re-
maining dihedral angle around the C_7—C_8 bond has been incremented some 25°
to minimize corresponding nitrogen and oxygen interatomic distances [Fig. 3(c)].
The interatomic distances in the overlap of prenalterol and isoprenaline are shown
in Figure 3(d). The main obvious difficulty is to reduce the beta- and *para*-hydroxy
interatomic distance within the constraints of the phenoxypropanolamine con-
former. Hydrogen-bond projections to receptor atoms permit a tolerance in this

distance, but further rotation around the C_7—O_3 bond much further than $30°$ is precluded with a planar—OCH_2 moiety. Forcing the anisole moiety out of plane within the agonist conformation by fixing the —CH_2 moiety at $60°$ to the plane of the aromatic ring might, therefore, raise the degree of agonist binding at the expense of increased deformation of the receptor for both agonist and antagonist binding. The predicted agonist conformer of a benzdioxepine molecule (Fig. 7) shows the intermolecular tolerances in comparative overlap with the isoprenaline conformer to reduce to ~ 0.5 Å. The benzdioxepine molecule has the highest intrinsic stimulatory activity of any phenoxypropanolamine agent (refer to Table I) and is in accord with the proposed agonist form for prenalterol.

The deformational analysis shows that within the model of agonist binding of potent phenoxypropanolamine conformers the steric freedom of the *ortho*-substituent is largely maintained but slight deformations in the neighboring 1-oxygen and 3-hydrogen atom positions occur on tighter binding. Deformation in the position of the aromatic ring 5-hydrogen atom in position a_2 [Fig. 1(b)] is not detectable and has a constant effect, if present, for the set of derivatives examined. The findings are consistent with the proposed contracted agonist conformer within the activated β_1-adrenoceptor complex. The hydrogen-bond distances in the activated form of the proton transfer system using the contracted phenoxypropanolamine conformer are shown in Figure 8(a). The relative positions of the residues compared with those for the isoprenaline conformer are given in Figure 8(b).

6. Conclusions

A small but important effect on the control of stimulant activity by *ortho*-substituents in β_1-adrenoceptor agents of the phenoxypropanolamine type can be attributed to the enthalpy associated with deformational effects on contraction of a given binding conformer. The deformations are dependent on the constraints of the receptor site that require a planar phenoxy moiety. Minimal basis-set calculations were found consistent with more extensive calculations to give insight into the underlying causes of "in-plane" σ-bond deformations. The defined contracted agonist conformer is of sufficient accuracy for calculating observed entropic differences between binding of agonist and antagonist conformers in an explicit model of transmembrane proton transfer within β_1-adrenoceptor α-helices.

Bibliography

[1] R. H. Davies, Int. J. Quantum Chem., Quantum Biol. Symp. **14**, 221 (1987).
[2] D. Timms, A. J. Wilkinson, D. R. Kelly, K. J. Broadley, and R. H. Davies, Int. J. Quantum Chem., Quantum Biol. Symp. **19**, 197 (1992).
[3] D. Timms, A. J. Wilkinson, D. R. Kelly, K. J. Broadley, and R. H. Davies, Receptors and Channels **2**, 107 (1994).
[4] A. J. Wilkinson and R. H. Davies, Int. J. Quantum Chem., Quantum Biol. Symp. **18**, 151 (1991).
[5] M. Nikishawa, M. Kanno, H. Kuriki, H. Sugihara, M. Motohashi, K. Itoh, O. Miyashita, Y. Oka, and Y. Sanno, Life Sci. **16**, 305 (1975).
[6] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **4**, 187 (1983).

[7] C. D. Strader, I. S. Sigal, M. R. Candelore, E. Rands, W. S. Hill, and R. A. F. Dixon, J. Biol. Chem. **263**, 10267 (1988).

[8] R. H. Davies and L. H. Smith, Int. J. Quantum Chem., Quantum Biol. Symp. **7**, 331 (1980).

[9] R. Clarkson and R. Dowell, personal communication.

[10] R. H. Davies, Int. J. Quantum Chem., Quantum Biol. Symp. **4**, 413 (1977).

[11] D. E. Leahy, P. J. Taylor, and A. R. Wait, Quantum Struct.-Act. Relat. **8**, 17 (1989).

[12] W. G. Richards, R. Clarkson, and C. R. Ganellin, Philos. Trans. Soc. Lond. B **272**, 75 (1975).

[13] M. N. Paddon-Row, C. Santiago, and K. N. Houk, J. Am. Chem. Soc. **102**, 6561 (1980).

[14] W. J. Hehre, L. Radom, and J. A. Pople, J. Am. Chem. Soc., **94**, 1496 (1972).

[15] G. M. Anderson III, P. A. Kollman, L. N. Domelsmith, and K. N. Houk, J. Am. Chem. Soc. **101**, 2344 (1979).

[16] M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Raghavachari, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol, and J. A. Pople, Gaussian 88 (Gaussian, Inc., Pittsburgh, PA).

[17] M. M. Nafissi-Varchei, D. P. Koharshi, M. S. Puar, and A. T. McPhail, J. Org. Chem. **42**, 3345 (1982).

[18] P. G. Jones, G. M. Sheldrick, R. Glenn, and A. J. Kirby, Z. Crystallogr. **163**, 85 (1983).

[19] C. M. Gramaciolli, R. Destro, and M. Simonetta, Act. Crystallogr. B **29**, 129 (1968).

[20] A. Pross, L. Radom, and N. V. Riggs, J. Am. Chem. Soc. **102**, 2253 (1980).

[21] J. W. Black and P. Leff, Proc. R. Soc. Lond. B **222**, 141 (1983).

[22] R. H. Davies, S. C. Sykes, and K. J. Broadley, unpublished work.

[23] R. H. Davies, R. D. Bagnall, B. Crooks, and W. G. M. Jones, Int. J. Quantum Chem. **13**, 53 (1986).

PNA Complexes of Polynucleotides and Polyamides: Structure of Two- and Three-Stranded Chimeric Helices Revealed by Conformational Analysis

I. A. IL'ICHEVA, E. V. TEREKHOVA, S. Yu. TSYBENKO,
A. G. KUNITZIN, AND V. L. FLORENTIEV

*V.A. Engelhardt Institute of Molecular Biology, Russian Academy of Sciences,
Vavilov Str. 32 Moscow B-334, 117984 Russia*

Abstract

We have performed a conformational analysis of double-stranded (dA:pT)$_5$ and triple-stranded (dA:pT $\cdot$ pT)$_5$ helices for all possible variants of mutual orientation of oligoamide and oligonucleotide strands by means of AMBER 3.0. Computation results showed that the conformational flexibility of chimeric helices is practically like the DNA flexibility, although orientation of atoms around the amide bond is almost planar. *cis-* and *trans*-orientations are close in energy. Permissible changes in helical parameters of chimeric helices practically coincide with the corresponding parameters of double- and triple-stranded DNA helices. Double-stranded chimeric helices exhibit a tendency to twist accompanied by helical pitch decreasing. Three-stranded chimeric complexes, on the contrary, exhibit a tendency to unwinding. Energy gain of chimeric helices is noticeable. Thus, double-stranded chimeras are characterized by the energy of 20 kcal/mol per monomer unit lower than double-stranded DNAs. The energy gain of triple-stranded chimeric complexes is about 40 kcal/mol per monomer unit. There is qualitative correlation between the experimentally obtained enthalpy of chimeric complexes and their calculated potential energy. It fully explained the ability of oligoamides to interact with DNA following oligoamide strand invasion of the duplex through D-loop formation. The dependence of energy on mutual strand orientation in chimeric duplexes is weak. Energy penalty of duplexes with parallel orientation of 5′ → 3′ and N → C chain vectors is about 0,7 kcal/mol per monomer unit. The dependence of energy on mutual strand orientation in chimeric triplexes is much more appreciable. The most advantageous is parallel orientation of 5′ → 3′ and N → C vectors of Watson–Crick chains accompanied by antiparallel orientation of the Hoogsteen oligoamide chain. It was shown that the stability of double-stranded oligonucleotides may be increased as a result of oligoamide insert of three or four monomer units in one of the oligonucleotide chains. The length and base sequence in the insert allowed one to modulate the degree of duplex stabilization. It is important that such stabilization may be obtained without any distortion in vector character of nucleotide duplex formation. It is evident that this method of stabilization of helices is suitable also for triplexes. Moreover, in this way, one can overcome the difficulties connected with the low penetration ability of PNA in living cells. © 1994 John Wiley & Sons, Inc.

Introduction

Three years ago, Nielsen and his colleagues designed a polyamide that could recognize both double- and single-stranded DNAs through Watson–Crick and Hoogsteen base pairing [1,2]. The structural element of this polyamide was the 2-aminoethylglycine unit and thymine attached through methylenecarbonyl group. Their complexes with oligodeoxyadenylic acid were studied in detail [1–5]. They

are, as a rule, three-stranded, consisting of one oligodeoxynucleotide and two oligoamide chains. The possibility of duplex formation consisting of thymine-constituting oligoamide and oligodeoxyadenylic acid was also registered [1], but rarely. If the oligoamide chain carries not only thymines, but also cytosines, the stoichiometry of complexes 2:1 is preserved [6]. But if all four nucleic bases are present in the oligoamide chain, then only duplexes were registered [7]. Melting temperatures of chimeric complexes are considerably higher than those for double- or triple-stranded helical DNAs of the same sequence. Binding of oligoamide chains to DNA may influence the biological functioning. Thus, the complete inhibition of restriction enzyme cleavage was obtained as a result of PNA binding [4]. There is also an indication of violations of transcription and translation as well as a gene-specific antisense effect in mammalian cells [8,9] in consequence of PNA formation.

But it is useful to investigate another possible utilization of the discovered oligoamides. We propose to use a short PNA insert in a single oligonucleotide chain as it may influence the stability of their duplexes with intact complementary oligonucleotides. The character of the insert (its length and base sequence) are the parameters that may allow us to modulate the stability of double-stranded oligonucleotides.

Conformational possibilities of chimeric helical complexes are not sufficiently known. There are two amide groups in any monomer unit of the oligoamide chain, and some peculiarities of conformational behavior of chimeric duplexes and triplexes may be expected. The vector character of oligoamide and oligonucleotide chains noticeably enlarges the number of principle different conformers. Hence, conformational possibilities of PNAs demand spatial investigation.

In the light of all the above-stated, it is very important to have available information about conformations of PNA. AMBER 3.0 [10] was used for the complete conformational analysis of double-stranded $(dA:pT)_5$ and triple-stranded $(dA:pT \cdot pT)_5$ helices. We also analyzed the influence of short PNA inserts in one strand of double-stranded decanucleotides $(dA:dT)_{10}$ (both in the dA chain, and in the dT chain) on helical parameters and their alteration on the boundary of the insert and estimated the possibility of duplex stabilization with the help of such inserts.

Models of Helical Chimeric Complexes: Starting Structures for Energy Minimization

There is convincing evidence that proved the existence of Watson–Crick and Hoogsteen binding between thymines of oligoamides and adenines in oligonucleotide chains in PNA duplexes and triplexes (Fig. 1). If we take into consideration the isomorphism of amide and nucleotide monomer units, we may use as starting coordinates of double- and triple-stranded DNAs those obtained from fiber diffraction data of Arnott and his colleagues [11,12]. It should be emphasized that each of Arnott's structures allowed us to obtain four variants of starting structures of oligoamides (Fig. 2). Structures of the I and III types differ from the structures of II and IV types in the chain direction.
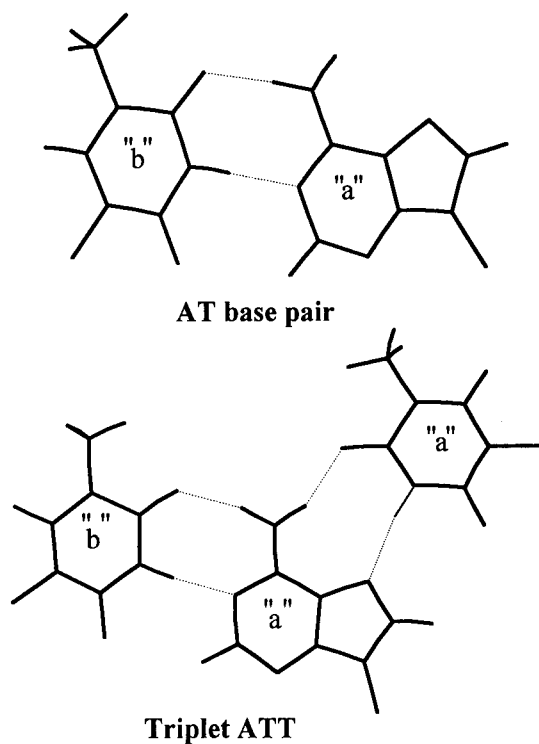
AT base pair

Triplet ATT

Figure 1.   Hydrogen bonding in the A:T base pair and in the A:T·T base triplet. "a" and "b" denote the nucleobase side in accordance with [14].

(a) Models of Double-Stranded Chimeric Structures

There are two principal different chimeric duplexes: with parallel and antiparallel orientation of $5' \rightarrow 3'$ and $N \rightarrow C$ vectors of Watson–Crick chains [Fig. 3(A)]. Each of them could be realized by two different schemes: I-st and III-d for antiparallel and II-nd and IV-th for parallel orientation of the chains. If we use X-ray coordinates of both the B- and A-forms of DNA, we dispose of four starting points for antiparallel and four starting points for parallel structures for energy optimization of chimeric duplexes.

(b) Models of Double-Stranded Oligonucleotides with Oligoamide Insert

Oligoamide inserts of 2, 3, and 4 monomer units were executed, in turn, in every chain of double-stranded decanucleotides $(dA:dT)_{10}$. Calculations were done for structures with inserts both in dA- and dT-chains. Figure 4 represents the molecular structure on the boundaries of the inserts. All variants of the mutual orientations of the chains in the insert and in the decanucleotide were investigated [Fig. 3(B)].

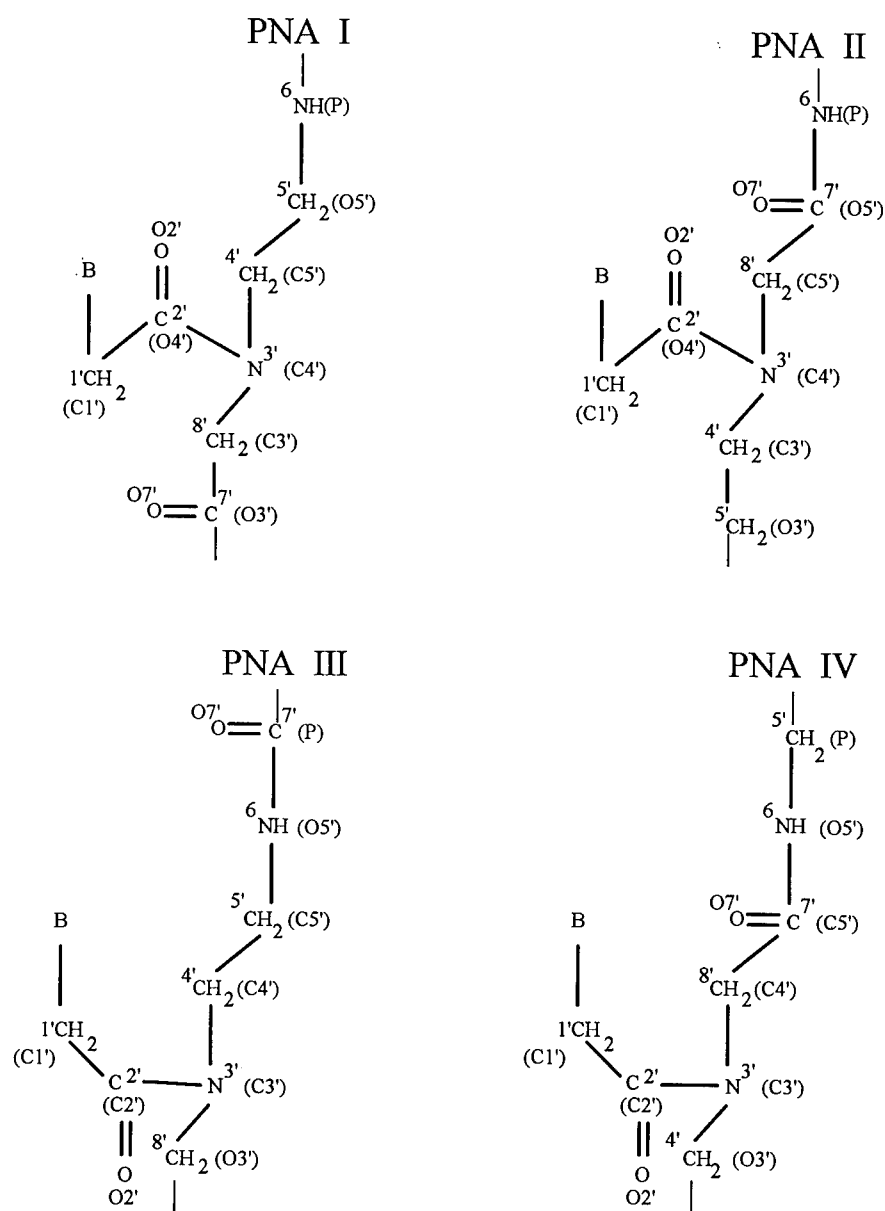## PNA I

PNA II

PNA III

PNA IV

Figure 2.   Four types of atomic disposition in the oligoamide chain isomorphic to DNA.
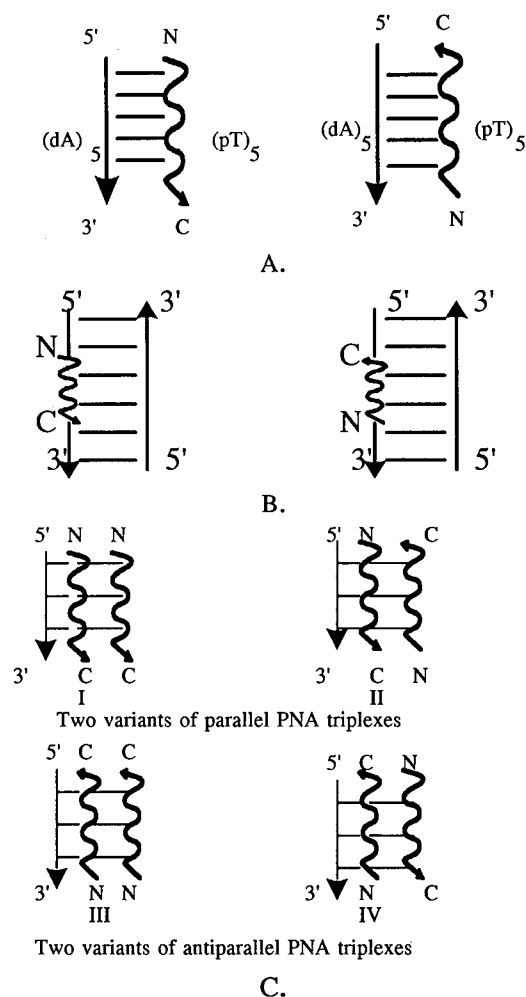
Figure 3.   Different variants of chain orientations in chimeric complexes: (A) two variants of chain orientations in PNA duplexes; (B) two variants of orientations of PNA insert in DNA duplexes; (C) four variants of chain orientations in PNA triplexes.

(c) Models of Triple-Stranded Structures

In triple-stranded structures, every type of mutual orientation of Watson–Crick chains allowed, in addition, two variants of Hoogsteen chain orientation [Fig. 3(C)]. Hence, there are four essentially different chimeric triplexes. Each of these variants may be realized with the help of the four different starting schemes presented in Figure 2. Therefore, we dispose of 16 different starting points for energy optimization of chimeric triplexes.
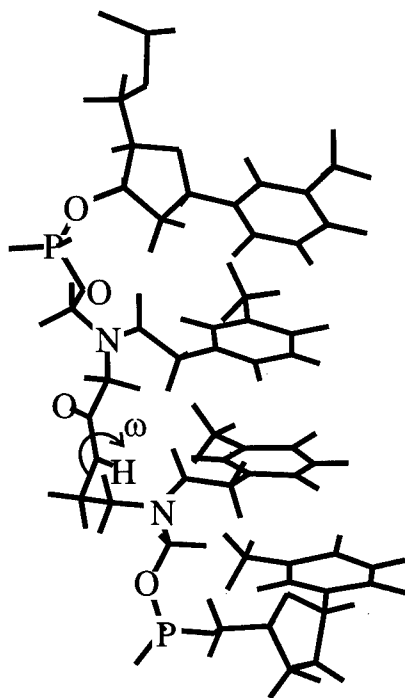
Figure 4. Peculiarities of molecular structure on the boundaries of the PNA insert in DNA duplexes.

Computational Method

(a) Scanning of Helical Parameters Space

We have written some auxiliary programs that allowed us to obtain atomic co-ordinates of double- and triple-stranded regular helices from coordinates of the first monomer unit. We obtained a conformational map for every type of different structures described above by the two-step procedure. As the first step, we searched for the local minima of the structure, responding to the fixed helical parameters (d, τ) scanned in the wide, but reasonable values, for the mutual orientation of bases in base pairs (or triplets). It is natural that in the result of the arbitrary helical transformation of monomer coordinates the bond lengths and bond angles in the sections of oligomers, connecting monomer units, are far from equilibrium. But several cycles of energy optimization are enough to bring them near to equilibrium. A further optimization process brings structure to the local minima of potential energy with barely fixed helical and base parameters. The next step was energy optimization without limitation on the mutual base orientation, which brings the structure to the local minima that is maximally close to fixed in the first step. Conduction of this procedure in all reasonable regions of the helical parameters (d, τ) allowed us to judge the potential surface of every type of chimeric structure.

(b) Scanning of Dihedral Angles' Space

It was shown earlier by conformational computations of canonical antiparallel [13] and parallel structures of polynucleotides [14] that there are two main stable and essentially different regions of conformational angles in DNA duplexes. The first is characterized by a (*gauche, gauche-*)-region of orientation around C4'—C5' and O5'—P bonds. It is the region of the lowest value of conformational energy and is typical for experimentally defined duplex structures of different forms of DNA, RNA, and DNA triplexes. The second is characterized by a (*trans, trans*)-region of orientation around these bonds. We called it the MIN2 region. It is uncommon and experimentally realized in only a few cases. One of examples is conformation of dT chain in hybrid duplexes poly(A):poly(dT) [15]. We propose that this conformation may play a significant role in the structures when hydrogen bonds in double-stranded complexes differ from the canonical Watson–Crick type, namely, in parallel stranded duplexes.

The amide angle (ω) is typical for the polyamide chain. Orientation around the C $\rightarrow$ N bond is nearly planar, owing to its partial double character. All *cis*- and *trans*-orientations are possible. In polypeptides, *trans*-orientation around this bond is favored, due to steric repulsion of C^{α} atoms of the neighboring residues (Pro is an exception to the general rule). We had to elucidate to what extent the frozen rotation around this bond influences the conformational mobility of the whole chimeric duplexes and triplexes. For this purpose, we had to obtain conformational maps by the way described above, using four essentially different conformations of monomer unit (γ, α, ω): (*gauche, gauche-, trans*), (*gauche, gauche-, cis*), (*trans, trans, trans*), and (*trans, trans, cis*).

(c) Methods of Optimization

All structures were minimized in two stages: At first, we used the Steepest Descent algorithm with a force criterion 0.5 kcal per step (or 1000 max. cycles). It was followed by the algorithm of Polak–Ribierre (Conjugate Gradients) with a force criterion of 0.01 kcal per step.

(d) Electrostatic Energy Term

There are no charges for the oligoamide unit in the AMBER 3.0 program. We had to add them into the program and coordinate their magnitude with the whole parameter system. For this purpose, we calculated atomic charges of the oligoamide unit as well as charges for the nucleotide unit by two different semiempirical methods—CNDO and AM1 and compared them with the charges accepted for peptides and nucleotides in AMBER 3.0. The differences in atomic charges obtained by the AM1 method and charges accepted in the AMBER 3.0 program are strongly pronounced, but they are less pronounced in the case of CNDO charges. In the AMBER 3.0 program, the value of the H-bonding energy as well as the torsion energy depends on the magnitude of the atomic charges. Therefore, we had to choose the values that are closer to the inner AMBER charges—charges obtained by the CNDO method.
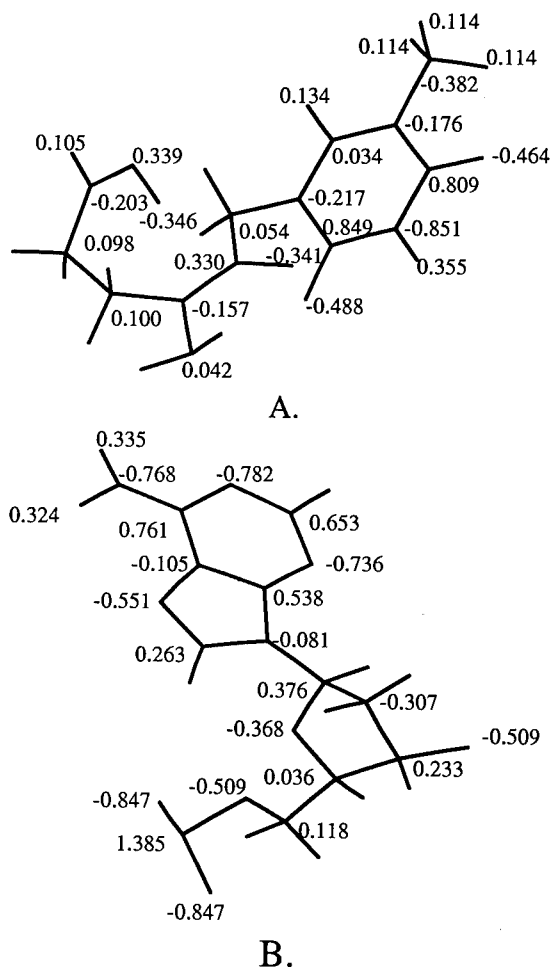
Figure 5. Atomic charges in monomer units: (A) oligoamide unit; (B) oligonucleotide unit.

The resulting set of atomic charges of monomer units is presented in Figure 5. Calculations were held down by the distant dependent dielectric permittivity $\varepsilon = R_{ij}$ (scale factor = 1).

Results and Discussion

(a) Double-Stranded Complexes of PNA

Conformational and helical parameters and corresponding potential energy per monomer unit in double-stranded complexes of PNA are presented in Table I for each of four conformational regions in both variants of mutual chain orientations. For a more convenient comparison, we provided the optimization of double-

TABLE I. Conformation and intramolecular potential energy of chimeric two-stranded complexes (dA)$_5$:(pT)$_5$.

Chain orientation	Conformational region	Oligo(dA) chain								W–C oligo(pT) chain	Helix parameters		Energy[a] (kcal/mol)
		χ	α	β	γ	ϵ	ζ	P	Θ_m	ω chain	d (Å)	τ (°)	E_{tot}
5'→3' C←N	tgg	137.3	283.6	178.5	62.6	183.4	247.1	160.3	39.4	188.4	2.7	39.6	−80.4
	cgg	134.1	283.4	172.6	65.6	187.6	231.9	169.6	39.8	353.5	2.94	46.2	−78.8
	ttt	156.2	102.2	184.9	177.8	275.7	288.7	158.0	45.1	185.6	2.91	32.2	−76.5
	ctt	123.6	160.6	172.2	172.5	181.3	257.1	178.9	37.7	352.9	2.86	48.7	−79.2
5'→3' N→C	tgg	129.2	289.5	164.5	69.6	184.9	231.5	167.6	40.5	172.1	2.66	45.3	−76.2
	cgg	142.7	279.3	180.5	61.6	186.8	252.5	160.3	41.0	1.0	2.90	42.8	−77.1
	ttt	126.7	152.8	177.0	173.4	183.5	263.4	165.9	42.0	185.6	2.86	44.9	−79.7
	ctt	124.4	152.7	174.6	175.0	180.8	267.3	170.3	38.5	1.8	2.88	45.5	−76.6
5'→3' 3'←5'	(dA)$_5$:	96.2	287.2	173.3	65.7	187.0	274.2	80.2	41.7		3.16	35.3	−62.5
	(dT)$_t$[b]	131.0	289.8	163.1	61.6	180.1	261.4	163.3	34.2				
5'→3' 5'←5'	(dAT)$_5$:	93.0	282.9	165.8	61.1	181.1	272.3	74.6	41.7		3.17	31.0	−57.7
	(dTA)$_5$	117.7	288.6	167.4	63.5	188.2	273.6	110.5	42.1				

[a] The energy value is given per nucleotide pair.
[b] For all duplex DNA, the parameters of the oligo(dA) chain are shown in the first line, the parameters of Watson–Crick oligo(dT) chain are shown in the second line.

stranded pentadeoxynucleotides $(dA:dT)_5$ and $(dAT:dTA)_5$ in the same parametrization and optimization regime.

It is obvious from the data in Table I that double-stranded helices twisted and the helical pitch became smaller if one changes the thymine-carrying oligonucleotide chain to the oligoamide chain. It resembles DNA twisting in high salt solutions and formation of the C-form of DNA [16]. It is obvious that at the physical bottom of such twisting is the lack of high negative charges in the polyamide chain. The most typical limits of PNA twisting are 40°–46°. But more twisted ($\tau = 49°$, for [cis, trans, trans]-type of conformer) so considerable more untwined ($\tau = 32°$, for [trans, trans, trans]-type of conformer) PNA conformations are possible. It is important to note that these considerable changes in conformation characterize PNA duplexes with antiparallel orientation of chains vectors. Just antiparallel is the lowest potential energy duplex, namely, the (trans, gauche, gauche)-type of conformer. Hence, conformational computations indicate that the most probable chimeric duplex is characterized by antiparallel orientation of the 5' → 3' and N → C vectors of Watson–Crick chains. It has many advantages both in potential energy and in conformational entropy. The best conformer of this kind is presented in Figure 6. It is interesting that in the experimental work of Nielsen et al. [7] it was shown that if the stochiometry 1:1 is strictly fulfilled (in the cases when all four nucleic bases presented in PNA structure) exactly antiparallel orientation of the strands is observed.

The limits of conformational angles changing in PNA duplexes of the (gauche, gauche)-type are identical to their limits in oligonucleotides with the exception of glycoside angle x. The last accept only values typical for the B- but not for the A-form of DNA.

The atomic orientation around the amide bond is nearly planar. Its deviation from planarity does not exceed 8°. But there is no noticeable definition of conformational flexibility of chimeric helices.
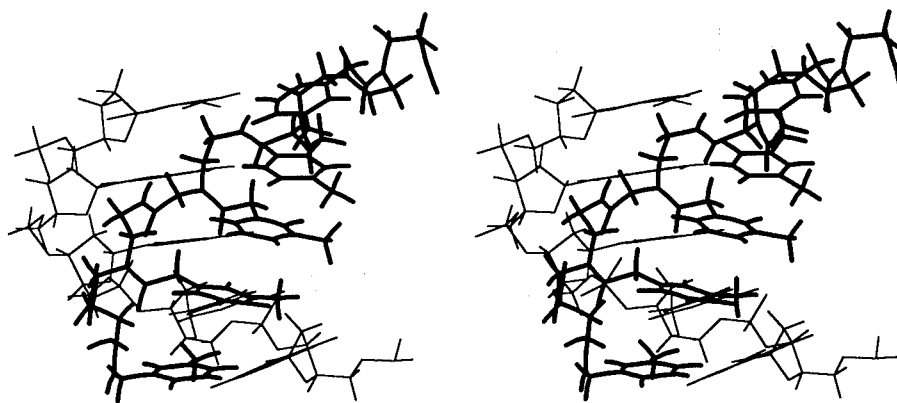

Figure 6.   Conformer with antiparallel orientations of W–C strands in chimeric duplexes.
Stereoview.

Intramolecular potential energy of hybrid complexes is lower than those of oligonucleotide duplexes. The energy gain is about 20 kcal/mol per nucleotide pair.

(b) Double-Stranded Oliginucleotide Complexes with PNA Insert in One of the Chains

The most important thing to know before planning experimental work is to what extent the short PNA insert may disturb the double-stranded helix. The energy gain due to the type of insert is also informative. Therefore, in Table II, we present the local helical parameters at the center of the inserts and at its boundaries for every chain direction of the inserts in $(dA)_{10}$ and in $(dT)_{10}$ chains of the duplex. Only conformations with the lowest intramolecular energy are presented. From the data obtained, it is evident that antiparallel orientation of the oligoamide chain in the insert gives the most stable decamers. But if the length of the insert is less than 3 monomer units, it causes an energy penalty in comparison with the homooligonucleotide. Thus, the intramolecular potential energy of $(dA:dT)_{10}$ is about -625 kcal/mol, and the most stable form of decamer with an antiparallel insert in the $(dA)_{10}$ chain is only about -607 kcal/mol. But if the length of the insert in the $(dA)_{10}$ chain exceeds 3 monomer units [for inserts in the $(dT)_{10}$ chain, 4 monomer units), it caused a noticeable energy gain. It reaches the values of about -30 kcal/

TABLE II. Perturbation of double-stranded DNA with the short inserts of PNA.

Type of insert in decamers		Local parameters of nonhomogeneous helices						Energy[a] (kcal/mol)
		$D_{5'}$ (Å)	$\tau_{5'}$ (°)	d_{pna} (Å)	τ_{pna} (°)	$d_{3'}$ (Å)	$\tau_{3'}$ (°)	
$(pT)_2$	$5' \rightarrow 3'$ $C \leftarrow N$	3.58	33.5	3.24	30.0	2.95	31.3	-595.3
	$5' \rightarrow 3'$ $N \rightarrow C$	3.44	35.2	3.12	31.6	3.27	32.6	-593.0
$(pT)_3$	$5' \rightarrow 3'$ $C \leftarrow N$	3.35	36.7	3.43	31.9	2.88	32.8	-619.9
$(pT)_4$	$5' \rightarrow 3'$ $C \leftarrow N$	3.33	37.1	3.24	34.1	3.00	34.5	-643.3
$(pA)_2$	$5' \rightarrow 3'$ $C \leftarrow N$	3.10	28.4	2.74	28.2	3.42	27.5	-607.2
	$5' \rightarrow 3'$ $N \rightarrow C$	2.91	24.7	2.42	24.3	3.06	31.9	-597.9
$(pA)_3$	$5' \rightarrow 3'$ $C \leftarrow N$	3.83	32.2	3.05	26.7	3.08	27.0	-637.4
$(pA)_4$	$5' \rightarrow 3'$ $C \leftarrow N$	3.9	33.5	3.07	26.7	2.99	27.2	-655.5

[a] The energy value is given per double-stranded decamer.

mol for inserts in the $(dA)_{10}$ chain, or less—about -18 kcal/mol for inserts in the $(dT)_{10}$ chain.

There is no twisting of the helices possessing short PNA inserts. Thus, the helical twist in the 3'-end and the center of the insert is about 34° and slightly rises to the 5'-end of the $(dT)_{10}$ chain. Even more unwounded helices are stabilized due to the inserts in the $(dA)_{10}$ chain. There is unhomogeneity of helices, possessed inserts, and it is more pronounced if these inserts are in the $(dT)_{10}$ chain. But this unhomogeneousness caused only weak distortions of the helical axis, as it is comparable with unhomogeneousness due to sequence changing. The most stable structures of decamers with 4 monomer unit inserts in $(dA)_{10}$ and $(dT)_{10}$ chains are presented in Figure 7.

(c) Triple-Stranded Complexes of PNA

We classify triple-stranded helical complexes of PNA on the basis of mutual orientation of the $5' \rightarrow 3'$ and $N \rightarrow C$ vectors of the Watson–Crick chains. If this orientation is parallel, we will call them **parallel**. In the reverse case, we also call them **antiparallel**. Each of these types of triplexes contain two subtypes, depending on mutual orientation of the Hoogsteen pair of chains.

Conformational and helical parameters of the pentamers $(dA:pT \cdot pT)_5$ with the lowest intramolecular potential energy for each of four types of triplexes, mentioned above, are presented in Table III. For a more convenient comparison, we present the results of the optimization of pentadeoxynucleotides $(dA:dT \cdot dT)_5$ in the same parametrization and optimization regime from two different variants of starting points: molecular structure with C3'-endo [12] and C2'-endo [17] sugars. The results of the computations show that there is now conformational differences between chimeric and nucleotide triplexes. C3'-endo-sugar is the preferable conformation in all kinds of triplexes. An exception is the conformation of the completely parallel triplex. The most stable conformation of the helix of this kind is characterized with the C2'-endo form of the sugar ring. It should be noted that the helical twist of this form is higher, and its pitch considerably lower, than in the others. Moreover, it is an excellent illustration of the coupling of conformational parameters in helices, because the glycoside angle X noticeably rises in accordance with the changing of the sugar form.

Conformers with the parallel type of chain orientation in chimeric triplexes are characterized with the lowest potential energy. The best conformer of this kind is presented in Figure 8. This type of triplex allows also the largest limits of the alteration of helical and conformational parameters, which is evidence of its largest conformational entropy. So, it may be supposed that the **parallel** type of triplexes is the most probable in experimental conditions. The exact preference for this type of mutual chain orientation in hybrid complexes of $(dA)_{10}$ and the C-lyzine, N-acridine derivative of $(pT)_{10}$ was defined by Nielsen et al. in [1]. But the stoichiometry of the complexes was not known definitely in this work. In the next work of this group of authors [2], the stoichiometry 2 PNA:1 DNA was established by UV-titration curves.
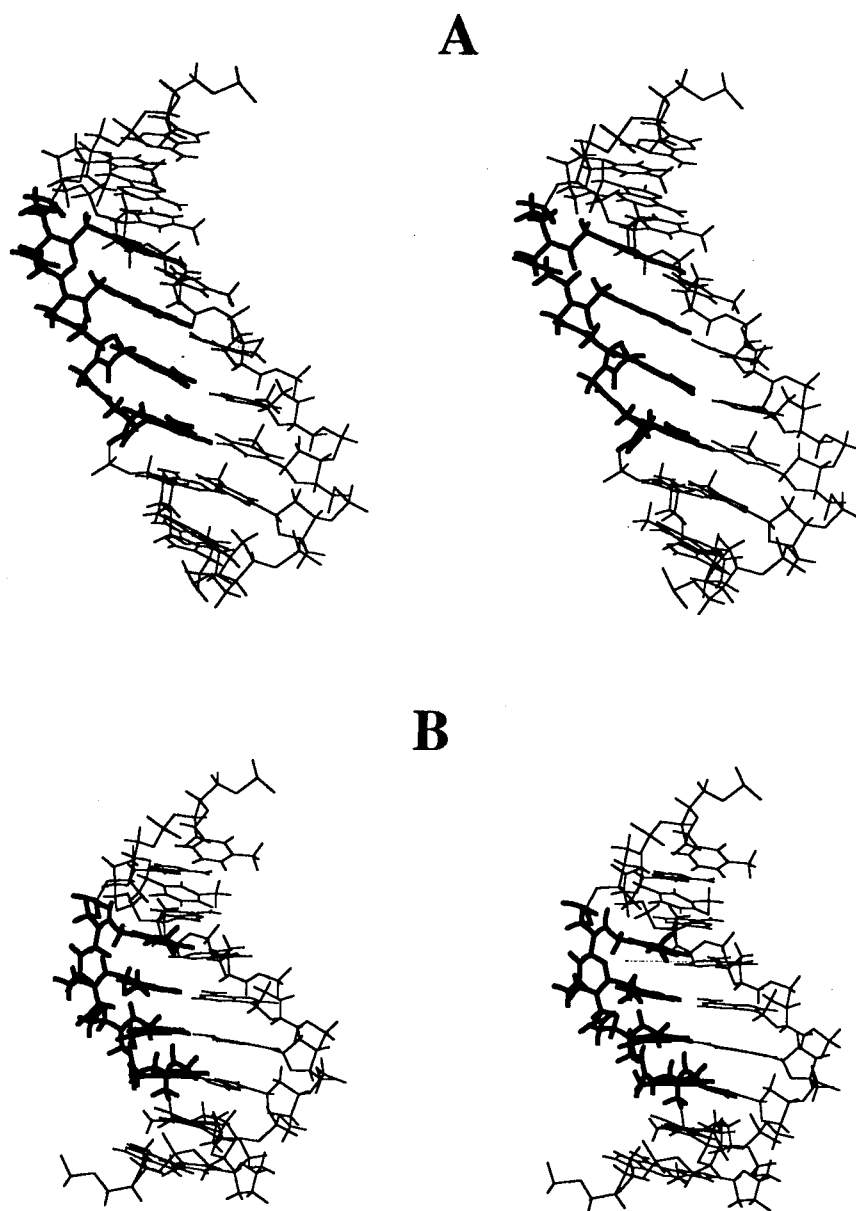
A

B

Figure 7.   Conformers with PNA inserts in the DNA duplexes: (A) insert of $(pA)_4$ in the duplex $(dA)_{10} \cdot (dT)_{10}$; (B) insert of $(pT)_4$ in the duplex $(dA)_{10} \cdot (dT)_{10}$.

TABLE III. Conformation and intramolecular potential energy of chimeric three-stranded complexes (dA)₅·(pT)₅·(pT)₅.

| | Conformational angles (in degrees) | | | | | | | | | | Helix parameters | | Energy[a] |
| | Oligo(dA) chain | | | | | | | | W–C oligo(pT) chain | Hoogst oligo(pT) chain | | | |
Chain orientation	χ	α	β	γ	ζ	ε	P	Θ_m	ω	ω	d (Å)	τ (°)	E_{tot} (kcal/mol)
5'→3' C→N / N→C	85.9	286.4	173.4	68.1	293.9	204.7	16.7	40.6	351.5	350.	3.32	22.4	−128.8
5'→3' N→C / C→N	86.7	279.6	170.3	69.5	294.6	189.0	17.9	40.1	348.5	179.4	3.51	25.1	−131.0
5'→3 C→N / C→N	84.3	279.5	169.0	69.4	296.2	190.2	18.2	40.0	179.9	347.1	3.44	27.1	−127.7
5'→3' N→C / N→C	121.1	291.0	184.1	59.9	268.2	188.6	154.1	36.0	180.5	342.1	2.72	36.1	−128.7
(dA)₅:[b]	109.9	293.2	170.0	64.6	269.8	186.8	108.7	44.1					
(dT)₅·	82.9	285.2	175.0	66.3	290.8	189.1	23.5	39.4			3.7	28.1	−87.5 [12]
(dT)₅	84.2	283.4	175.2	64.4	294.6	193.6	12.4	40.4					
(dA)₅:[b]	109.2	292.0	170.6	64.1	271.9	188.2	110.7	44.1					
(dT)₅·	104.1	286.5	172.5	61.2	274.5	188.8	101.7	42.1			3.7	27.3	−83.9 [17]
(dT)₅	86.8	279.4	170.7	58.5	286.9	185.9	46.9	42.7					

[a] The energy value is given per nucleotide pair.

[b] For all triplex DNA, the parameters of the oligo(dA) chain are shown in the first line, the parameters of Watson–Crick oligo(dT) chain are shown in the second line, and the parameters of the Hoogsteen oligo(dT) chain are shown in the third line.
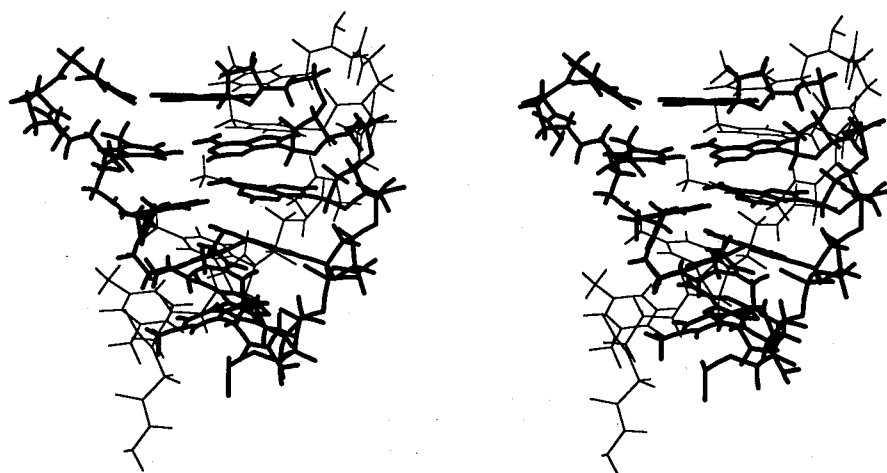
Figure 8.   Conformer with parallel orientations of W–C strands in chimeric triplexes.
Stereoview.

The intramolecular potential energy in chimeric triple-stranded complexes is considerably lower than in oligonucleotide triplexes. The energy gain is about 40 kcal/mol per monomer unit. So, it is not surprising that oligoamides interact with DNA, forming PNA triplexes in the process of D-loop formation [3–5]. Almarsson et al. [18] provided molecular mechanics calculations of such a complicated system, comprising two oligoamide chains and double-stranded DNA, forming a D-loop. They indicated two main reasons stabilizing such structures: "van der Waals attraction between relatively nonpolar peptide strands, and stabilising electrostatic effect arising from removal of one phosphodiester strand." Our calculations of three-stranded forms of PNA, in turn, show their intrinsic stability.

Acknowledgments

Bibliography

[1] P. E. Nielsen, M. Egholm, R. H. Berg, and O. Buchardt, Science 254, 1497 (1991).
[2] M. Egholm, O. Buchardt, P. E. Nielsen, and R. H. Berg, J. Am. Chem. Soc. 114, 1895 (1992).
[3] D. Y. Cherny, B. P. Belotserkovskii, M. D. Frank-Kamenetskii, M. Egholm, O. Buchardt, R. H. Berg, and P. E. Nielsen, Proc. Natl. Acad. Sci. U.S.A. 90, 1667 (1993).
[4] P. E. Nielsen, M. Egholm, R. H. Berg, and O. Buchardt, Nucleic Acids Res. 21, 197 (1993).
[5] S. K. Kim, P. E. Nielsen, M. Egholm, O. Buchardt, R. H. Berg, and B. Norden, J. Am. Chem. Soc. 115, 6477 (1993).
[6] M. Egholm, P. E. Nielsen, O. Buchardt, and R. H. Berg, J. Am. Chem. Soc. 114, 9677 (1992).

[7] M. Egholm, O. Buchardt, L. Christensen, G. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden, and P. E. Nielsen, Nature **365**, 566 (1993).

[8] P. E. Nielsen, M. Egholm, R. H. Berg, and O. Buchardt, Anti-Cancer Drug Design **8**, 53 (1993).

[9] J. C. Harvey, Science **258**, 1481 (1992).

[10] U. C. Singh, P. K. Weiner, J. W. Caldwell, and P. A. Kollman, AMBER (UCSF), version 3.0, Department of Pharmaceutical Chemistry, University of California, San Francisco (1986).

[11] S. Arnott, P. J. C. Smith, and R. Chandrasekaran, *Handbook of Biochemistry and Molecular Biology,* 3rd ed., Nucleic Acids, Vol. II, G. D. Fasman, Ed., Cleveland, Ohio (1976), p. 411.

[12] S. Arnott and E. Selsing, J. Mol Biol. **88**, 509 (1974).

[13] I. A. Il'icheva, V. G. Tumanyan, A. E. Kister, and V. G. Dashevsky, Int. J. Quantum Chem. **17**, 321 (1980).

[14] I. A. Il'ychova, Yu. P. Lysov, A. A. Chernyi, A. K. Shchyolkina, B. P. Gottikh, and V. L. Florentiev, J. Biomol. Struct. Dynam. **7**, 879 (1990).

[15] S. B. Zimmerman, and B. H. Pheiffer, Proc. Natl. Acad. Sci. U.S.A. **78**, 78 (1981).

[16] S. Arnott and E. Selsing, J. Mol. Biol. **98**, 265 (1975).

[17] G. Raghunathan, H. T. Miles, and V. Sasisekharan, Biochemistry **32**, 455 (1993).

[18] O. Almarsson, Th. C. Bruice, J. Kerr, and R. N. Zuckermann, Proc. Natl. Acad. Sci. U.S.A. **90**, 7518 (1993).

Stability of the 1,3-Substituted 1,4-Dihydropyridines: Substituent Effects on the Acid Catalyzed Hydration and Oxidation Reactions

EMIL POP,[1,2]* MING-JU HUANG,[2] MARCUS E. BREWSTER,[1,2] AND NICHOLAS BODOR[1,2]

[1]Pharmos Corporation, 2 Innovation Drive, Alachua, Florida 32615 and [2]Center for Drug Discovery, College of Pharmacy, University of Florida, Gainesville, Florida 32610

Abstract

1,3-Substituted-1,4-dihydropyridines easily add water to the 5,6-double bond in acid catalysis resulting in 6-hydroxy-1,4,5,6-tetrahydropyridines. The influence of various substituents in position C-5 and C-6 over the hydration of 1-methyl-1,4-dihydromethylnicotinate, used as a model compound, was investigated in the framework of the AM1 molecular orbital approximation. Since the rate-limiting step of the reaction is a proton transfer from the acidic species to the C-5 position of the substrate, calculated proton affinities (PA) were used as reactivity indexes. The results, in agreement with experimental evidences, indicated that electron-donating (+I) substituents increase the PA and destabilize 1,4-dihydropyridines towards hydration, while electron-withdrawing (−I) groups have the opposite effect. Calculated vertical ionization potentials (Ip) indicate that similarly +I groups facilitate the one-electron oxidation, while −I groups stabilize the molecules toward this reaction. Several molecular properties derived from the principle of maximum hardness were also used for the investigation of the stability of the dihydropyridines. © 1994 John Wiley & Sons, Inc.

Introduction

Dihydropyridines have an important role in numerous vital biological processes being constituents of the ubiquitous NADH ↔ NAD$^+$ coenzyme system; accordingly the chemistry of these compounds has been extensively investigated [1,2]. Recently, 1,4-dihydropyridine ↔ pyridinium quaternary salt moieties have been used in redox targetor based chemical delivery systems (CDSs) to transport drugs specifically to the central nervous system [3–5]. While of remarkable efficiency in intravenous formulations, CDSs are less suitable for oral administration due to the instability of the 1,4-dihydropyridines in the harsh acidic condition of the stomach. At low pH these derivatives easily add water across the 5,6-double bond resulting in 6-hydroxy-1,4,5,6-tetrahydropyridines [6,7]. This unwanted transformation represents a practical inconvenience, since the product of hydration, both reduces the lipophilicity of the CDS and prohibits the *in vivo* oxidation to the quaternary pyridinium salt, two major requirements of this approach. Modalities to circumvent this problem

---

* To whom correspondence should be addressed.

CCC 0360-8832/94/010173-08

have been recently explored [8]. The possibility to increase the stability of the 5,6-double bond toward the acid catalyzed hydration by the structural manipulation of the dihydropyridine moiety is examined theoretically herein. The stability of dihydropyridines toward the chemical oxidation to pyridinium salts is also investigated.

1-Methyl-1,4-dihydromethylnicotinate (I, Fig. 1) was selected as a model compound for the study. The influence of substituents at C-5 and C-6 positions (R_1 and R_2, respectively) on the reactivity of I towards acid catalyzed hydration and oxidation reactions was examined.
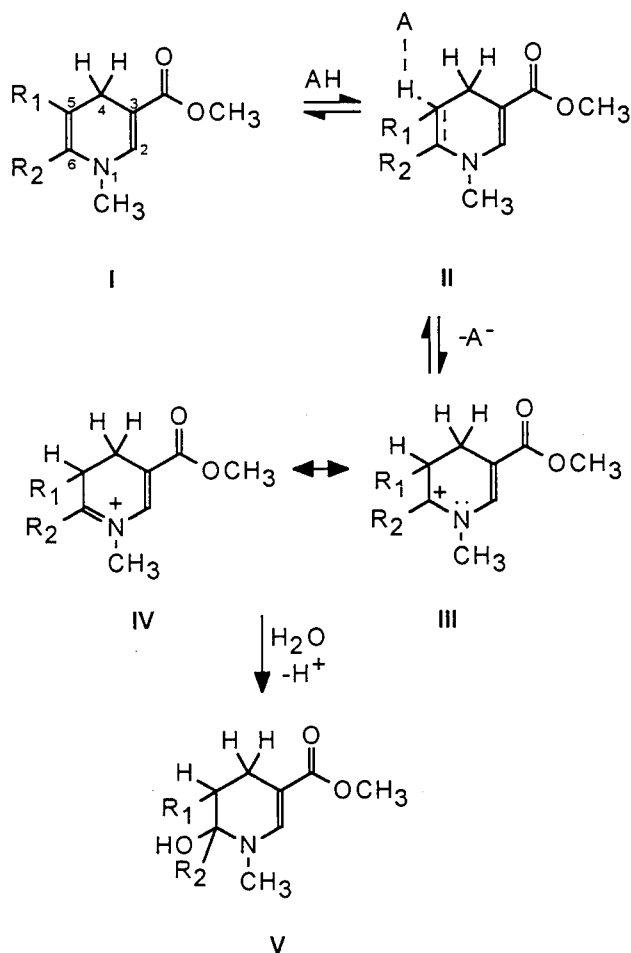
R$_1$ and R$_2$: see Table I

Figure 1.   The mechanism of acid catalyzed hydration of 1,4-dihydropyridines.

The theoretical study was performed in the framework of the AM1 molecular orbital approximation [9], based on its well documented ability to accurately predict heats of formation and molecular geometries. No solvent effects were considered in this study.

Methods

Theoretical studies were performed using the AM1 molecular orbital method [9,10] included in the MOPAC (version 5.10) package. A Tektronix Computer Aided Chemistry (CAChe™) Worksystem run on an Apple Macintosh™ II computer was used for all computations. The structural input was generated using a Macintosh interface and all starting geometries were found by using molecular modeling (MM2) to optimize the geometries. The Broyden–Fletcher–Goldfarb–Shanno method [11–14] was used to optimize geometries as a function of the total molecular energy. All geometric variables were optimized. The dynamic "level-shift" method [15] was used to improve the convergence of the self-consistent field (SCF). The "precise" option was implemented for tightening the convergence criteria for all optimizations. The closed- and open-shell species were investigated using the restricted Hartree–Fock (RHF) approach.

Results and Discussion

The mechanism of the acid catalyzed hydration of the 1,4-dihydropyridines [16,17] is illustrated in Figure 1, for the case of 1-methyl-1,4-dihydromethylnicotinate (I). In the first, rate determining step of the reaction, a proton transfer from the acid species AH to the C-5 position [the β-position of the enamine system N(1)-C(6)-C(5)] of the substrate occurs. The most probable transition state of this step is II; the proton then forms a bond with C-5 by converting the π pair of electrons in a σ pair as in any electrophilic addition. The resulting intermediate II has a positive charge on C-6 which can be accommodated by the N-1 atom, in the canonical form IV. The combination of the highly reactive intermediates III ↔ IV with a species carrying an electron pair, such as OH⁻ in the case of hydration is then very rapid, the final product being the 6-hydroxy-1,4,5,6-tetrahydro derivative V. Obviously, other nucleophiles if present in the reaction media can also attack the positive ion.

The influence of several electron-donating (+I) (methyl) and electron- withdrawing (−I) (halogen) substituents at C-5 (R_1) and/or C-6 (R_2) over the hydration reaction was investigated. It is known [18] that generally electron withdrawing groups in α reduce the reactivity of double bonds presumably by decreasing the electronic density, electron donating groups having the opposite effects.

However, a better explanation of the phenomenon can be given by using a thermodynamic criterion. Since formation of the intermediate carbanion III ↔ IV is the rate limiting step of the reaction, it is reasonable to consider calculated proton affinities (PA) of I as reactivity indexes. PA of a compound B to form the conjugate acid HB⁺ [19] can be determined by using the equation:

$$\text{PA}(B) = \Delta H_\text{f}(H^+) + \Delta H_\text{f}(B) - \Delta H_\text{f}(HB^+)$$

which becomes for the case described in Figure 1:

$$\text{PA}(\mathbf{I}) = \Delta H_\text{f}(H^+) + \Delta H_\text{f}(\mathbf{I}) - \Delta H_\text{f}(\mathbf{III})$$

where ΔH_f are calculated heats of formation for **I** and **III** (**III** and **IV**, being canonical forms have the same ΔH_f). Since AM1 gives a very poor estimate of the heat of formation of H^+ (calculated: 314.9; observed 367.2 kcal/mol [20]), the experimental value was used in calculating PA.

The heats of formation for **I**, **III**, and **V** as well as the PA of **I**, for various R_1 and R_2 are presented in Table I. The results indicate that PA is indeed influenced by the nature and position of the substituents. There is a 14.5 kcal/mol difference between the lowest and highest calculated PA. Electron-withdrawing substituents at C-5, C-6, or both positions decrease the PA and stabilize **I** toward hydration. A decrease of the PA with the decrease of the electronegativity (relative to H) was registered for the R_2 substituents (at C-6) (order of PA: F > Cl > Br > I). In the case of the R_1 substituents (at C-5) PA increased with the decrease of the electronegativity (order: I > Br > Cl > F). When both C-5 and C-6 protons bear an electronegative substituent (compounds **I-10** and **I-11**) PA had the lowest value. Based on these considerations the most stable combinations toward hydration are, in order **I-10** (5,6-difluoro), **I-6**, and **I-11** (5-fluoro and 5,6-dichloro, respectively). There is a 10 kcal/mol difference between the PA of **I-1** (the unsubstituted compound) and that of **I-10** (the 5,6-difluoro derivative). Electron-donor substituents (methyl) at C-6 (**I-12**) or at both C-5 and C-6 (**I-14**) increase the PA and consequently the hydration reactivity (4 kcal/mol difference), while the substituent at

TABLE I.   Calculated AM1 heats of formation (ΔH_f) and proton affinities (PA) (kcal/mol).

| Compound | Substituents | | ΔH_f (kcal/mol) | | | PA (kcal/mol) |
	R_1	R_2	**I**	**III**	**V**	**I**
1	H	H	−54.1	100.7	−121.5	212.5
2	H	F	−95.0	62.7	−169.8	209.5
3	H	Cl	−59.1	98.7	−124.4	209.4
4	H	Br	−46.8	112.4	−106.4	208.0
5	H	I	−35.4	124.6	−95.0	207.2
6	F	H	−98.9	64.6	−167.9	203.7
7	Cl	H	−61.1	101.2	−127.9	204.9
8	Br	H	−49.6	113.2	−114.2	204.5
9	I	H	−38.3	123.6	−101.6	205.3
10	F	F	−138.0	−27.3	−211.7	202.0
11	Cl	Cl	−64.5	99.0	−126.3	203.7
12	H	CH₃	−60.1	90.6	−121.3	216.5
13	CH₃	H	−62.1	95.0	−129.0	210.2
14	CH₃	CH₃	−65.3	85.8	−123.9	216.5

C-6 (**I-13**) decreases somewhat the PA (with 2.3 kcal/mol). All these findings are in agreement with available experimental evidences [16].

Another important transformation of the 1,4-dihydropyridines is the oxidation to the quaternary salts (**VIII**, Fig. 2). While the *in vivo* oxidation is a requirement of the CDS approach, a fast chemical oxidation can be disadvantageous for the drug (handling, storage, and formulation are difficult). The chemical oxidation was well investigated in a series of recent studies [21–23].

It was shown, for example [22], that observed second order rate constants of ferricyanide-mediated oxidations well correlated ($r = 0.96$) with AM1-derived vertical ionization potentials; these results were consistent with a mechanism of oxidation involving an initial rate-determining (or partially rate-determining) electron loss from the heterocycle, followed by sequential proton and electron transfers (Fig. 2). Calculated vertical ionization potentials (Ip) for various **I** (**1-14**) are presented in Table II. The results indicate that electron withdrawing groups (−I) increase Ip and stabilize dihydropyridines toward oxidation, while electron donating groups (+I) decrease the Ip. These data are in agreement with experimental findings [24] which indicate that +I groups (methyl) destabilize dihydropyridines, facilitating their oxidation (by single electron transfer). However, differences in calculated Ip are smaller (maximum 0.425 kcal/mol) as compared to differences in calculated PAs.

Several molecular properties derived from the principle of maximum hardness (PMH) [25,26] were applied to the dihydropyridines **I** in order to discuss their reactivity. The principle of maximum hardness is a result of the application of the
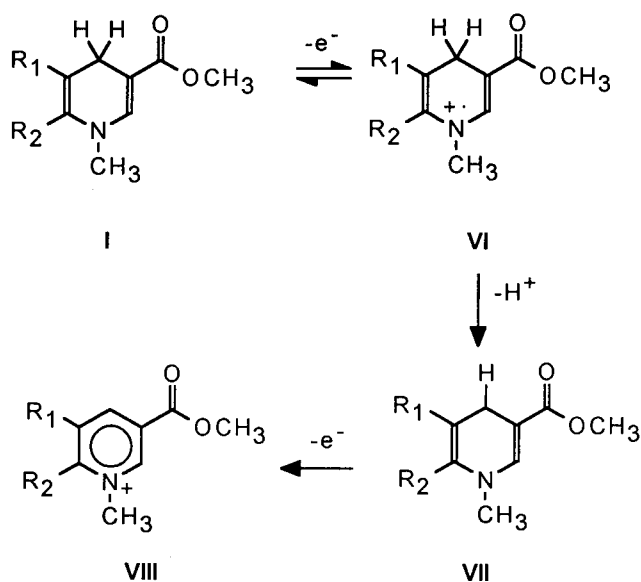

Figure 2.   Oxidation of 1,4-dihydropyridines.

TABLE II. Calculated vertical ionization potentials (Ip)
for I.

| Compound | Substituents | | Ip (ev) |
	R_1	R_2	
1	H	H	8.172
2	H	F	8.436
3	H	Cl	8.384
4	H	Br	8.396
5	H	I	8.384
6	F	H	8.303
7	Cl	H	8.323
8	Br	H	8.390
9	I	H	8.422
10	F	F	8.546
11	Cl	Cl	8.485
12	H	CH_3	8.100
13	CH_3	H	8.074
14	CH_3	CH_3	8.011

density functional theory (DFT) to chemistry. Hardness (η) is the HOMO–LUMO energy gap and can be defined by the equations:

$$\eta = (I - A)/2 = (E_{LUMO} - E_{HOMO})/2$$

where I is the ionization potential, A the electron affinity, E_{LUMO} and E_{HOMO} the energies of the lowest unoccupied and highest occupied molecular orbitals, respectively. The maximum hardness principle asserts that systems tend to be as hard as possible; thus a hard molecule has a large energy gap, and a soft molecule has a small gap. Bigger η means larger I and smaller A which implies that the system has a lower tendency to accept and to give away particles; that is the system is stable. Soft molecules (softness is defined by $\sigma = 1/\eta$) are more reactive than hard molecules. Another molecular property derived from DFT is the absolute electronegativity (χ) defined by the equation:

$$\chi = -\mu = (I + A)/2$$

(μ being the electronic chemical potential). χ is similar but not equal to the Mulliken electronegativity.

Calculated η, σ, and χ are collected in Table III. The results indicate that there are no significant differences in hardness and softness for various substituted dihydropyridines, which means that the reactivity of the dihydropyridines I1-14 is not much different. Small differences in hardness and softness for isomers and closely related compounds are not unexpected, however [26]. The results suggest that the proton addition is more dependent on the thermodynamic stability of the intermediate cation III than on the reactivity of I. Calculated electronegativity (χ)

TABLE III. Calculated highest occupied orbital (HOMO), lowest unoccupied molecular orbital (LUMO) energies (E), hardness (η), softness (σ), and absolute electronegativity (χ) (eV) for I.

Compounds	E_{HOMO}	E_{LUMO}	η	σ	χ
1	−8.172	0.182	4.177	0.239	3.995
2	−8.436	−0.067	4.185	0.239	4.252
3	−8.384	−0.038	4.173	0.240	4.211
4	−8.396	−0.058	4.169	0.240	4.227
5	−8.384	−0.176	4.104	0.244	4.280
6	−8.303	−0.079	4.112	0.243	4.191
7	−8.323	−0.065	4.129	0.242	4.194
8	−8.390	−0.119	4.136	0.242	4.255
9	−8.422	−0.117	4.153	0.241	4.270
10	−8.546	−0.332	4.107	0.244	4.439
11	−8.485	−0.262	4.112	0.243	4.374
12	−8.100	0.216	4.158	0.241	3.942
13	−8.074	0.186	4.130	0.242	3.944
14	−8.011	0.226	4.119	0.243	3.893

can be correlated to the rate of oxidation. A decreasing electronegativity indicates an increase of the rate of oxidation since lower electronegativity reflects a greater tendency to lose electrons. Data presented in Table III indicate, that $+I$ groups decrease χ, increasing the rate of oxidation, while $-I$ groups have the opposite effect. These data are in agreement with the previous findings based on calculated ionization potentials.

In summary, the AM1 study indicates that electron-withdrawing groups at C-5 and C-6 positions stabilize the 1,4-dihydropyridine moieties toward both acid catalyzed hydration and oxidation reactions. These results could have practical applications.

Acknowledgment

Bibliography

[1] U. Eisner and J. Kuthan, Chem. Rev. 72, 1 (1972).
[2] D. M. Stout and A. I. Meyers, Chem. Rev. 82, 223 (1982).
[3] N. Bodor, Drugs Future 6, 165 (1981).
[4] N. Bodor, Ann. N.Y. Acad. Sci. 507, 289 (1981).
[5] E. Pop, M. E. Brewster, and N. Bodor, Drugs Future 15, 473 (1990).
[6] A. G. Anderson, Jr. and G. Berkelhammer, J. Am. Chem. Soc. 80, 992 (1958).
[7] K. S. Choi and G. A. Alivisitos, Biochemistry 7, 190 (1968).
[8] E. Pop, Z. Z. Liu, J. Vlasak, W. Anderson, M. E. Brewster, and N. Bodor, Drug Targeting Delivery, in press.
[9] M. J. S. Dewar and D. M. Storch, J. Am. Chem. Soc. 107, 3898 (1985).
[10] M. J. S. Dewar, E. G. Zoebisch, E. I. Healy, and J. J. Stewart, J. Am. Chem. Soc. 107, 3902 (1985).

[11] C. G. Broyden, J. Inst. Math. Appl. **6**, 222 (1970).

[12] R. Fletcher, Comput. J. **13**, 317 (1970).

[13] D. F. Shanno, Math. Comput. **24**, 647 (1970); D. F. Shanno, J. Optim. Theory Appl. **46**, 87 (1985).

[14] D. Goldfarb, Math. Comput. **24**, 23 (1970).

[15] A. V. Mitin, J. Comput. Chem. **9**, 107 (1980).

[16] S. L. Johnson and P. T. Tuazon, Biochemistry **16**, 1175 (1977).

[17] F. Jordan, J. Am. Chem. Soc. **96**, 825 (1974).

[18] P. Y. Sollenberger and R. B. Martin, J. Am. Chem. Soc. **92**, 4261 (1970).

[19] M. J. S. Dewar and K. M. Dieter, J. Am. Chem. Soc. **108**, 8075 (1986).

[20] D. R. Stull and J. Prophet, JANAF Thermochemical Tables; NSRDS-NBS 37, 1971.

[21] M. Brewster, A. Simay, K. Czako, D. Winwood, H. Farag, and N. Bodor, J. Org. Chem. **54**, 3721 (1989).

[22] M. Brewster, J. Kaminski, Z. Gabanyi, K. Czako, A. Simay, and N. Bodor, Tetrahedron **45**, 4395 (1989).

[23] M. E. Brewster, M-J Huang, J. J. Kaminski, E. Pop, and N. Bodor, J. Comput. Chem. **12**, 1278 (1991).

[24] E. Pop, E. Shek, T. Murakami, and N. Bodor, J. Pharm. Sci. **78**, 609 (1989).

[25] R. G. Parr and Z. Zhou, Acc. Chem. Res. **26**, 256 (1993).

[26] R. G. Pearson, Acc. Chem. Res. **26**, 260 (1993).

Quantitative Structure–Inhibitory Activity Relationships of Substituted Phenols on *Bacillus Subtilis* Spore Germination

MING-JU HUANG AND NICHOLAS S. BODOR

Center for Drug Discovery, P.O. Box 100497, Health Science Center, Gainesville, Florida 32610-0497

Abstract

Our semiempirical method of calculating log P, which uses a linear combination of various combinations of descriptors based on the AM1-optimized geometry of the molecule, has been applied to a set of 38 substituted phenols and has been shown to be successful in providing a better quantitative structure–activity relationship (QSAR) than that of the Hansch-type approach in a study of the inhibitory activity of substituted phenols on *Bacillus subtilis* spore germination. This model shows that the calculated partition coefficient, the geometrical descriptors, and electronic effects are the major factors determining the biological activity. © 1994 John Wiley & Sons, Inc.

Introduction

The goal of workers in the area of quantitative structure–activity relationships (QSAR) has been the development of quantitative methods of determining the activities of a series of compounds. There have been numerous mathematical attempts to correlate molecular structure with drug activity. A significant aspect of these kind of studies is the often-found relationship between biological properties of molecules and their partition coefficient (log P).

Yasuda et al. [1] correlated directly the log P values of substituted phenols to their inhibitory activity (log $1/I_{50}$) on *Bacillus subtilis* spore germination. The log P values used were combinations of the experimental data and calculated results obtained by the Hansch fragment method. Klopman et al. [2] correlated this inhibitory activity with functional group descriptors in the same way that they developed their log P model. In this study, we used the calculated log P values from our BLOGP program [3]. By applying the same semiempirical method as before [4–8] to the entire set of 38 substituted phenols, we found that this method provides a better quantitative structure–activity relationship than that of the Hansch-type approach and also uses fewer parameters than does Klopman's approach in the study of inhibitory activity of substituted phenols on *B. subtilis* spore germination.

Methods

The entire set of 38 compounds were calculated by the Tektronix CAChe (Computer Assisted Chemistry) workstation. For each compound, MM2 [9] is the starting

point for the AM1 [10] calculation. In all cases, the default (Broyden–Fletcher–Goldfarb–Shanno) geometry search method was employed to yield full geometry optimization. The default SHIFT = 15 eV option was used to allow 15 eV of damping of SCF iterations to be determined by the rate of convergence, and the PRECISE option was used to strengthen the convergence criteria for the SCF procedure and for the geometry optimization. Based on the AM1-optimized geometry and the van der Waals radii of each atom, the molecular surface, volume, and ovality were calculated the same way as before [4–6].

The inhibitory effect of various concentrations of phenols on the germination rate in 0.1 mM L-alanine at pH 7.2 was studied by Yasuda et al. [1]. The molar concentration of a phenol necessary to cause 50% inhibition of the germination rate (I_{50}) was determined by them also. Each I_{50} value represents the mean of three determinations.

The log P value for each molecule was calculated by our BLOGP program on a microvax computer, and linear combinations of the calculated descriptors were fitted to the observed inhibitory activity (log $1/I_{50}$) by using a regression program developed for the VAX computer. As mentioned by Klopman et al. [2], the general form of the QSAR is

$$\log_{10}(\text{act.}) = a + b(\log_{10} P) + c(\text{other descriptors}) \tag{1}$$

and

$$\log_{10} P = \sum_i n_i P_i, \tag{2}$$

where the P_i correspond to the different molecular properties that we have developed for log P. Those "other descriptors" are descriptors of which some may contribute to log P, but may have an additional contribution to activity.

Results and Discussion

Instead of using the combination of experimental data and the calculated results from the Hansch fragment method for the log P, we used the BLOGP program on our VAX computer to calculate log P. The calculated log P value for each compound is listed in Table I. We first attempted to correlate the $\log_{10} 1/I_{50}$ activity with the partition coefficient alone, which resulted in quite a good correlation:

$$\log_{10} 1/I_{50} = 1.531 + 0.6725 \log P \tag{3}$$

$$n = 38, \qquad r = 0.825, \qquad \text{s.d.} = 0.426, \qquad F = 76.582.$$

The regression analysis for 38 substituted phenols is described in Eq. (3), where n is the number of compounds submitted to the regression; r, the correlation coefficient; s.d., the standard deviation; and F, the overall statistical significance of the equation. Our calculated log P alone provides a better structure–activity relationship than that of Yasuda's study. Overall, our calculated partition coefficients are quite close to the set that has been used by Yasuda et al. [1], except for the 2,3,4,5,6-Cl$_5$ phenol and hexachlorophene. The experimental log P data for 2,3,4,5,6-Cl$_5$ phenol

TABLE I.   Inhibitory activity of substituted phenols on *B. subtilis* spore germination.

No.	Phenol substituent	$\log 1/I_{50}$ (exp)	$\log 1/I_{50}$ (ours)	$\log P$ (calcd)	Ovality	HARD (eV)
1	2-CH_3	2.89	2.98	1.702	1.276	9.367
2	2-C_2H_5	3.31	3.18	2.101	1.321	9.363
3	2-$CH(CH_3)_2$	3.57	3.39	2.484	1.362	9.386
4	2-$CH_2CH(CH_3)_2$	3.62	3.54	2.860	1.410	9.374
5	2-$C(CH_3)_3$	4.46	3.69	2.855	1.379	9.399
6	2-Cl	3.25	3.16	1.829	1.251	9.289
7	2-NO_2	2.50	2.74	1.903	1.273	8.726
8	2-C_6H_5	3.55	3.55	3.152	1.382	8.813
9	2-OCH_3	2.20	2.40	1.353	1.303	9.174
10	2-CHO	2.42	2.32	1.278	1.264	8.924
11	2-$COOCH_3$	3.00	2.41	1.715	1.348	8.961
12	2-NH_2	1.52	1.32	0.188	1.273	9.048
13	4-CH_3	2.44	2.88	1.705	1.293	9.307
14	4-C_2H_5	2.80	3.07	2.121	1.341	9.298
15	4-$CH(CH_3)_2$	3.17	3.32	2.555	1.390	9.363
16	4-$CH_2CH(CH_3)_2$	3.52	3.49	3.002	1.449	9.333
17	4-$C(CH_3)_3$	3.52	3.57	2.889	1.408	9.365
18	4-Cl	3.10	2.99	1.778	1.268	9.219
19	4-NO_2	2.17	2.66	1.723	1.299	9.006
20	4-C_6H_5	3.89	3.43	3.194	1.386	8.610
21	4-CHO	1.70	2.19	1.171	1.290	9.046
22	4-$COOCH_3$	2.70	2.36	1.616	1.366	9.139
23	4-H	2.82	2.92	1.300	1.224	9.512
24	3-CH_3	2.70	2.93	1.682	1.292	9.410
25	2-$CH(CH_3)_2$, 5-CH_3	3.31	3.49	2.917	1.419	9.272
26	2,4,6-$(CH_3)_3$	2.85	3.20	2.597	1.390	9.122
27	2,4-Cl_2	3.62	3.29	2.346	1.294	9.026
28	2,4,5-Cl_3	3.89	3.44	2.773	1.325	8.833
29	2,4,6-Cl_3	3.89	3.49	2.802	1.327	8.889
30	2,3,4,6-Cl_4	3.57	3.75	3.276	1.344	8.702
31	2,3,4,5,6-Cl_5	3.46	3.97	3.639	1.357	8.596
32	(Hexachlorophene)	3.70	3.95	4.539	1.539	8.423
33	3-NO_2	3.05	2.49	1.703	1.300	8.799
35	2,5-$(NO_2)_2$	1.48	1.72	1.418	1.343	8.355
35	3-OH	2.42	2.24	0.851	1.259	9.370
36	3-OH, 5-CH_3	1.77	2.33	1.268	1.325	9.314
37	3-NH_2	1.35	1.18	0.046	1.275	9.053
38	4-Br	3.14	3.32	2.150	1.273	9.211

listed by Hansch and Leo [11] are 5.01, 5.12, 5.86, and 3.81. Our calculated log P for 2,3,4,5,6-Cl_5 phenol is 3.639, which is quite close to one of the experimental results. The experimental data for hexachlorophene listed by Hansch and Leo [11] are 2.62 and 7.54. The log P of 2.62 is at pH 12.5, and the compound forms a salt; the log P value of 7.54 is a calculated value and is based on the assumption of only mono ion partitions at pH 12.5, so neither of them is satisfactory for our purposes.
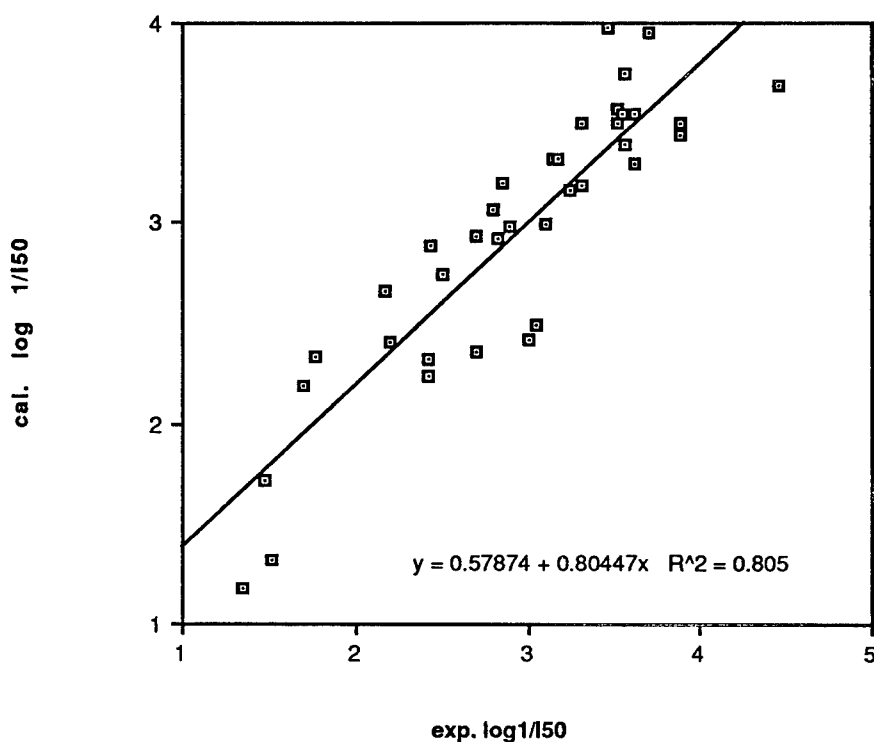
HUANG AND BODOR

exp. log1/I50

Figure 1.   The experimental values plotted against our calculated values.

The second most important parameter that correlated with the activity is the electronic descriptor HARD. The results are shown in Eq. (4):

$$\log_{10} 1/I_{50} = -4.9975 + 0.7388 \log P + 0.7029 \text{ HARD} \qquad (4)$$

$$n = 38, \qquad r = 0.867, \qquad \text{s.d.} = 0.375, \qquad F = 52.920,$$

where HARD is twice the absolute hardness of the molecule, as discussed by Pearson [12]: The absolute hardness is one-half of the calculated energy difference between the LUMO, the lowest unoccupied molecular orbital, and HOMO, the highest occupied molecular orbital, in units of eV. We know that soft molecules, with a small LUMO–HOMO gap, will be more polarizable than will hard molecules, with a large LUMO–HOMO gap. This suggests that electronic interactions, or removal of one electron from the oxygen in the phenol group or hydrogen bonds to a receptor from the phenolic group, play a significant role in the interaction of the receptor–ligand molecule.

The best relationship among the activity of *B. subtilis* spore germination and the hydrophobicity, the geometrical descriptors of the substituted phenols, and the electronic effects is given by Eq. (5):

$$\log_{10} 1/I_{50} = 0.1342 + 0.9631 \log P - 4.2143 O + 0.7036 \text{ HARD} \quad (5)$$

$$n = 38, \quad r = 0.897, \quad \text{s.d.} = 0.333, \quad F = 46.710,$$

where O is the ovality of the molecule and has been described in previous papers [4–6] and is dimensionless. Our standard deviation for this study, 0.333, is better than that of Klopman et al. [2], 0.43, by using three parameters instead of seven. From Eq. (5), we see that the partition coefficient and the electronic descriptor, HARD, have positive influences on the $\log_{10} 1/I_{50}$, and the geometric descriptor, ovality, has a negative influence on the $\log_{10} 1/I_{50}$.

Both the calculated and the experimental values of $\log_{10} 1/I_{50}$ from Eq. (5) are shown in Table I. It is fitted very well in all 38 substituted phenols. The 2-C(CH$_3$)$_3$ phenol has the largest residual of the 38 phenol substituents, 0.77, and it also shows a significant large residual in the study by Klopman et al. [2], 0.65, and by Yasuda et al. [1], 1.19. One may be quite suspicious about the experimental data of this 2-C(CH$_3$)$_3$ phenol because isobutyl phenol and t-butyl phenol have identical inhibitory activities, but for 2-CH$_2$CH(CH$_3$)$_2$ and 2-C(CH$_3$)$_3$ substituents, the two isomers have 0.84 differences in the $\log_{10} 1/I_{50}$ value.

In conclusion, we find that by using our semiempirical approach to this QSAR study our results are better than those of previous studies. Figure 1 shows the experimental values plotted against our calculated values.

Bibliography

[1] Y. Yasuda, K. Tochikubo, Y. Hachisuka, H. Tomida, and Y. Ikeda, J. Med. Chem. **25**, 315 (1982).
[2] G. Klopman, K. Namboodiri, and M. Schochet, J. Comput. Chem. **6**, 28 (1985).
[3] The BLOGP program in the VAX version.
[4] N. Bodor, Z. Gabanyi, and C.-K. Wong, J. Am. Chem. Soc. **111**, 3783 (1989).
[5] N. Bodor, and M.-J. Huang, J. Pharm. Sci. **81**, 272 (1992).
[6] N. Bodor and M.-J. Huang, J. Pharm. Sci. **81**, 954 (1992).
[7] G. Hochhaus, P. Druzgala, R. Hochhaus, M.-J. Huang, and N. Bodor, Drug Design Discov. **8**, 117 (1991).
[8] G. Kumar, M.-J. Huang, R. Hammer, and N. Bodor, J. Pharm. Sci. **83**, 117 (1994).
[9] J. T. Sprague, J. C. Tai, Y. Yuh, and N. Allinger, J. Comput. Chem. **8**, 581 (1987).
[10] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 (1985).
[11] C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology* (Wiley-Interscience, New York, 1979).
[12] R. G. Pearson, Proc. Natl. Acad. Sci. U.S.A. **83**, 8440 (1986).

The Investigation of Structural Transitions and Energy Transfer in DNA Solution in Range 4.2–273 K Using 6-Thioguanine

YU. V. RUBIN AND YU. P. BLAGOY

Institute for Low Temperature Physics and Engineering, Ukrainian Academy of Sciences, Kharkov, 310164, Ukraine

V. A. BOKOVOY

Institute of Molecular Biology, Russian Academy of Science, Moscow 117986, Russia

V. S. LEONTIEV AND R. G. LEE

Institute for Low Temperature Physics and Engineering, Ukrainian Academy of Sciences, Kharkov, 310164, Ukraine

Abstract

The electronic structure of the low-temperature phosphorescencent probe 6-thioguanine was investigated. Structural transitions in some alcohols (ethanol, glycerol, propanediol) were obtained using this probe in the range 4.2–273 K. Aqueous solutions of native and denatured DNA and those of native DNA with propanediol and DMSO added were studied in the range of 4.2–273 K. The analysis of the luminescence spectra of DNA solutions permits the assumption of the energy transfer to the probe on UV radiation of DNA. © 1994 John Wiley & Sons, Inc.

Introduction

Studies of low-temperature structural transitions in DNA solutions are of great interest to biophysics. To study such transitions, we investigated the structural changes in aqueous solutions of DNA in the absence and presence of alcohols. The studies were carried out in the range from 4.2 to 273 K using the phosphorescent probe method.

Earlier, the method of luminescence probes was tested on studying proteins and membranes [1,2]. In the present work, 6-thioguanine (6SG) or its nucleoside were used as phosphorescent probes. Their electronic structures were studied by us in detail earlier [3–5]. It is known that 6SG is an antitumor drug used for leicoses treatment [6].

Experiment

6-Thioguanine (6SG) and 6-thioguanosine (6SGR) were synthesized at the Institute of Organic Synthesis, Latvian Academy of Sciences. (6SGR is a fixed N9H tautomer of 6SG.) 6-Thiopurine–riboside was purchased from the Serva Co. The

alcohols (ethanol, glycerol, propanediol [PD]) and dimethyl sulfoxide (DMSO) were preliminary distilled and dried. DNA from *E. coli* was used, into which 6SG was introduced by the biosynthesis method [7]. The concentrations were 10^{-5} *M* 6SGR in alcohols, and 10^{-3} *M* DNA in solution. The relative 6SG concentration in DNA was 1 nucleotide with 6SG per 3000 or 600 major nucleotides. The DNA solution contained $0.015 M$ NaCl and $0.0015 M$ sodium citrate. To obtain denatured DNA, its initial solution was kept in the boiling water bath for 30 min and then cooled down to 4°C.

DNA films colored with 6-thiopurine–riboside were also studied. The spectral features of the latter are close to those of 6SG [8]. The preliminary dried film was kept in an atmosphere of KCl aqueous solution of the required concentration, which provided the necessary humidity (80%).

To prepare experimental samples, a special cuvette (volume 0.3 mL) with the prepared solution was quickly dipped into liquid nitrogen. Then, the sample was cooled down to the liquid helium temperature. The rate of cooling was 6–9 K/min at 4.2–77 K and that of heating was 0.7–1 K/min in the range 4.2–273 K. The temperatures were measured with a differential cuprum–constantan thermocouple.

The phosphorescence spectra and temperature dependence of the probe's phosphorescence intensity were obtained using the laboratory luminescence setup incorporating a nitrogen–helium cryostat and a microcomputer. The exciting light wavelength was 350 nm, the observation wavelength being 485 nm in the studies of the temperature dependencies of probe emission.

Results and Discussion

Earlier, we carried out theoretical (CNDO/S methods) and experimental (absorption and luminescence spectroscopy, circular dichroism) studies on the electronic structure of 6SG [3–5]. The results showed, in particular, that UV absorption spectra have an intense long wavelength band (Fig. 1) that is due to the intramolecular transfer of an electron from the sulfur atom to the pyrimidine ring. The luminescence spectrum shows essentially phosphorescence. The CNDO/S method was used to calculate the excited-state energies (Table I) and the elements of the electronic structure (atomic charges, bond orders, spin densities) for the ground and excited states of the neutral and ionic forms of 6SG. The atomic contribution of the sulfur atom to the resonance integral was taken to be 14 eV and the single-center electron–repulsion integral was put at 7 eV. These parameters were derived from the calculation for a large group of sulfur-containing compounds (the ionic and tautomeric forms of 4SU, 2SU, 2.4SU, and others) [5].

The comparison of full atom charges of guanine and 6SG shows their similarity in the ground state. This permits the N9H tautomer of 6SG to enter into the DNA. The changes in the full atom charge on the transitions to the $S_{\pi\pi^*}$, $T_{\pi\pi^*}$, and $n\pi^*$ electronic states occur mainly at the C6 and S10 atoms. The largest change in the bond orders occurs for the C=S bond. This suggests that excitation in the above states is localized at the C=S fragment, which makes it the most photoreactive. According to theoretical calculations, the dipole moment of the neutral 6SG mole-
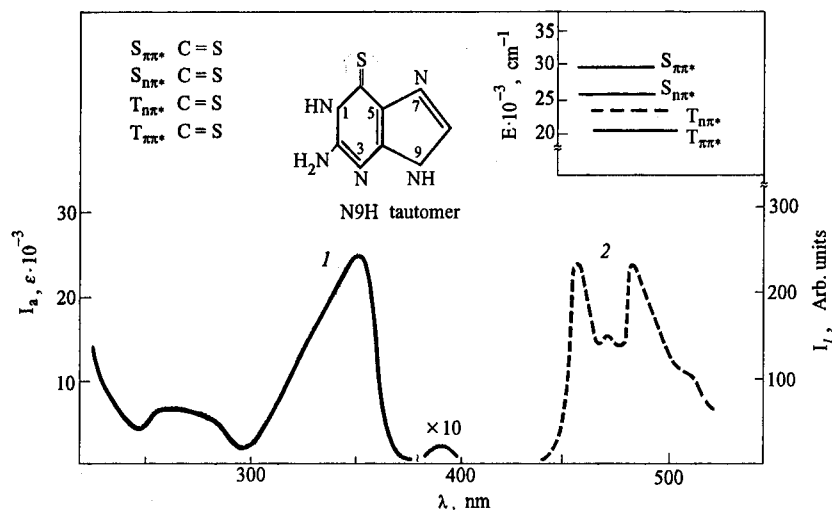
Figure 1. (1) Absorption and (2) phosphorescence spectra of the 6SG probe (N9H tautomer neutral molecule) in ethanol at 77 K. Top: Localization of excitation and the excited-state diagram.

cule decreases in the excited states: from 10.80 D in the ground state to 6.69 D in the $S_{\pi\pi^*}$ state and 6.67 D in the $T_{\pi\pi^*}$ state.

The analysis of spectroscopic and theoretical data permits the plotting of a diagram of the excited state of the neutral 6SG molecule (Fig. 1). The diagram shows, in particular, that 6SG can be a suitable phosphorescence probe to study structural changes and transitions in DNA and low molecular compounds in a wide temperature interval.

Using 6SG incorporated into DNA as a phosphorescent probe, we studied the temperature dependencies of the probe emission in water DNA solutions and in solutions with added PD and DMSO in a wide temperature interval. Since the

TABLE I. Calculated (Calc.) and experimental (Exp.) excited-states energies (E) and oscillator strengths (f) of electronic transitions of different ionic and tautomeric forms of 6SG.

Form of molecule	$E_{S_{\pi\pi^*}}$ (eV)		f	$E_{n\pi^*}$ (eV)	$E_{T_{\pi\pi^*}}$ (eV)	
	Calc.	Exp.			Calc.	Exp.
Cation	3.29	3.56	0.50	2.43	2.12	2.59
Neutral N9H	3.68	3.61	0.47	3.01	2.71	2.70
Neutral N7H	3.67	3.40	0.43	3.19	2.73	2.59
Anion N9H	4.00	3.92	0.33	3.73	3.23	2.87
Anion N7H	4.09	—	0.32	3.93	3.26	2.80
Dianion	3.73	3.87	0.36	3.39	2.93	2.80

temperature dependencies of the probe emission in DNA are complex in shape, we studied the temperature dependencies of our probe emission in simple systems: ethanol, glycerol, and PD.

The temperature dependencies of the 6SGR phosphorescence intensity, obtained on heating in glycerol, ethanol, and PD (Fig. 2), showed that the probe luminescence intensity changes with the changes in the aggregated states of alcohols [9–12]: on the devitrification, formation, and melting of crystals. The strong luminescence decrease (phosphorescence quenching) is observed in the range of 100–120 K in ethanol, in the range of 150–200 K for PD and its solutions, and in the range of 180–190 K for glycerol. The 6SGR luminescence also changes in ethanol at 125–150 K, which correlates with the formation and melting of ethanol crystals, as calorimetric data suggest [10,12]. The comparison between the calorimetric data for alcohols [9–12] and the temperature dependencies of the 6SGR emission intensity (Fig. 2) shows that the strongest decrease in the luminescence intensity (phosphorescence quenching) is observed when the liquid phase appears.

It is also seen in Figure 2 that in the range from 4.2 K to the temperature of the phosphorescence quenching onset the luminescence intensity of solid-state alcohols decreases 1.5–3 times in the ethanol–PD–glycerol row. This may be due to the slow processes of the thermoactivation mobility of the molecule fragments or the whole molecules of the matrix (glass) [13].

We also studied the temperature dependencies of the 6SGR phosphorescence intensity in the 15% water PD solution and in the 15% DMSO solution. In these experiments, phosphorescence quenching was observed in the ranges of devitrification of the amorphous phases for the solutions of the same concentration: in the range of 165 K for PD and in the range of 130 K for DMSO [9,14].

After this preliminary analysis of the probe emission thermograms in simple systems, we could examine the emission thermograms of the probe incorporated into DNA. As seen in Figure 3, the temperature dependencies of the 6SG phosphorescence in native DNA (nDNA) have specific features at 21, 64, 87, 140, 182,
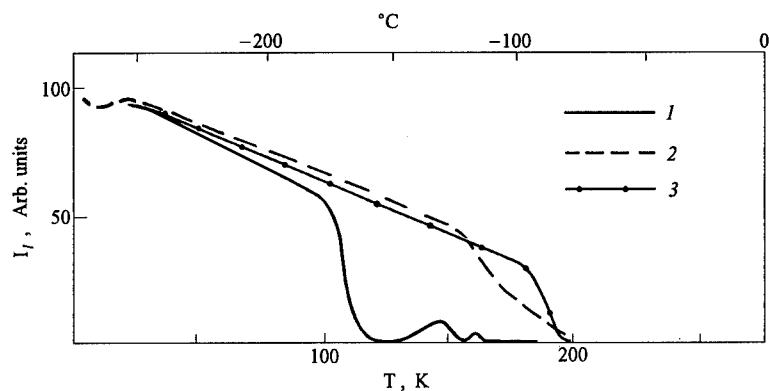

Figure 2.   The temperature dependencies of the 6SGR probe emission intensity: (1) in ethanol; (2) PD; (3) glycerol.
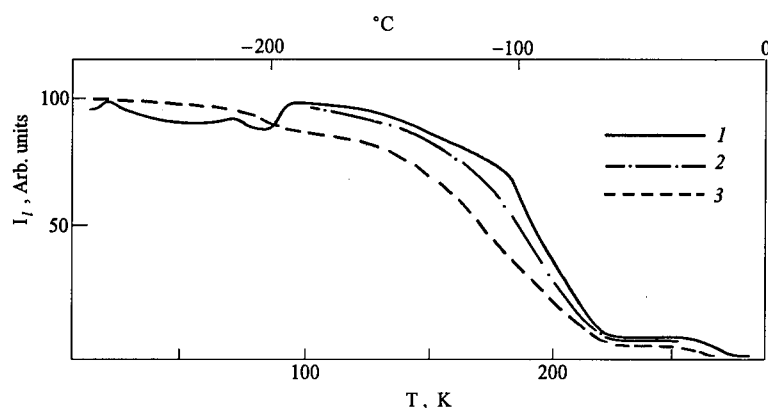
Figure 3.   The temperature dependencies of the emission of 6SG incorporated in DNA.
(1) Warming and (2) cooling of native DNA; (3) warming of denaturated DNA. Rates of
cooling/warming were 1 K/min.

and 268 K. The amplitude of the effect observed at 87 K decreases after the sample
annealing. The phosphorescence quenching of 6SG in DNA is observed at 182–
215 K. The temperature intervals of the of 6SG phosphorescence increase on cooling
of nDNA and of its phosphorescence quenching on heating are close. It is also seen
that the interval of the phosphorescence quenching of 6SG in denatured DNA shifts
toward lower temperatures as compared to that of nDNA. Unlike the nDNA having
peculiarities at 64, 87, 140, 182, and 268 K, the denatured DNA has them at 90,
140, and 268 K. The slope of the temperature dependence of the probe luminescence
in DNA at 21–140 K is significantly smaller than that of the corresponding curves
taken in alcohols (cf. Figs. 2 and 3).

Figure 4 shows that the 6SG phosphorescence quenching in DNA in 15% PD is
observed at lower temperatures (160–180 K). This temperature interval is char-
acteristic of devitrification of PD and its solution (cf. Fig. 2). The temperature
dependence of 6SG in DNA in the PD solutions manifests peculiarities at 9, 75,
and 160 K. At lower temperature, the probe phosphorescence quenching is observed
in the aqueous DNA solution with added DMSO. The range from 120 to 170 K is
close to the region of devitrification of the aqueous DMSO solutions [14]. The
preliminary studies on DNA films colored with 6-thiopurine–riboside (RH 80%)
show that the probe's phosphorescence quenching interval is shifted to the region
of higher temperatures (190–245 K).

The X-ray diffraction analysis of the data for the B-form DNA structure at 16
and 293 K [15,16] and the data for DNA with 6SG instead of G [6,7] allows the
conclusion that the $C{=}S$ fragment of the 6SG molecule, where the excitation is
localized, is within the major groove of DNA in aqueous solutions.

In DNA films, the probe and its $C{=}S$ fragment are located nearer to the phos-
phate groups or sugars than to the bases because the phosphates and sugars form
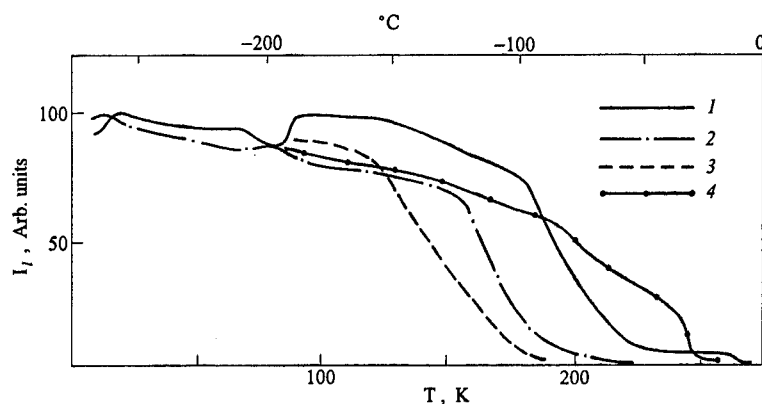about 80% of the surface of the DNA double helix [17]. Perhaps the above effect

Figure 4.   The temperature dependencies of the emission of 6SG incorporated into DNA
under different conditions: (1) water solution; (2) 15% PD solution; (3) 15% DMSO solution;
(4) hydrate film of DNA (probe 6-thiopurine–riboside was not incorporated into DNA,
but was placed in the DNA hydrate shell).

in DNA films explains the shift of the interval of the probe phosphorescence
quenching to the higher-temperature region because the binding energy of phos-
phates to water is known to be higher than that of the bases [17].

The analysis of data on DNA calorimetry [18,19] and the 6SGR phosphorescence
quenching (Fig. 2) in alcohols with an appearing liquid phase and the similarity
of the temperature intervals where the phosphorescence intensity increases or
quenches on cooling and heating of nDNA solutions allows the assumption that
the 6SG phosphorescence quenching in native DNA at 180–215 K (Fig. 3) is due
to the releasing of the brakes of the water molecule motions (devitrification) in the
major groove of DNA. This is also evidenced by the shift of the temperature interval
of the phosphorescence quenching with PD and DMSO added to the aqueous DNA
solutions.

Perhaps the increasing luminescence intensity of the aqueous nDNA solution at
87 K, which decreases on sample annealing, is due to the sample cracking. Earlier,
we observed the increase in the probe luminescence intensity on the cracking of
transparent alcohol glasses with a probe. The factors causing these peculiarities at
21, 64, and 140 K are being studied now.

We also studied the luminescence spectra of water DNA solutions with different
concentrations of the probe in DNA. The DNA with the probe concentration of 1
probe molecule per 3000 and 600 major nucleotides and DNA without the probe
(Fig. 5) were studied. The analysis of the luminescence spectra of the DNA solution
at λ_{ex} = 280 nm shows that a change from DNA without the 6SG probe to DNA
containing one 6SG residue per 3000 and 600 major nucleotides leads to a decrease
in the DNA fluorescence intensity at 330–380 nm and a simultaneous considerable
increase in the luminescence at 450–500 nm. The emission intensity in the range
450–500 nm appreciably exceeds the expected values of the 6SG emission and the
6SG concentration, and the efficiency of its absorption at λ_{ex} = 280 nm is taken
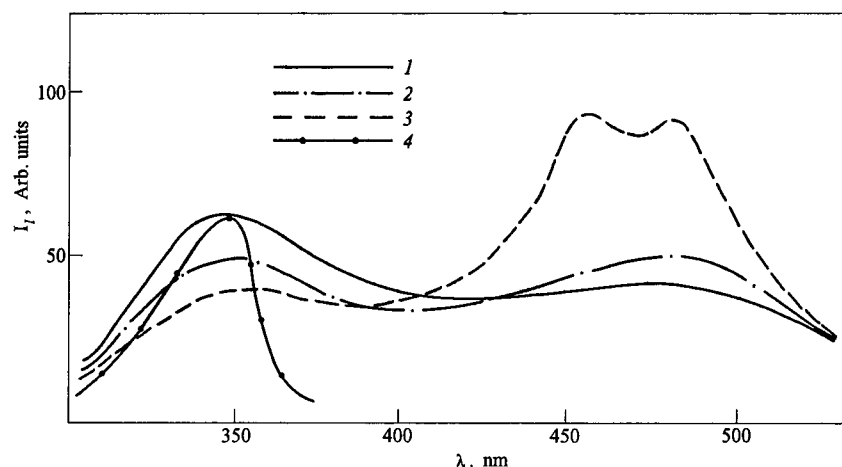
Figure 5. The luminescence spectra of DNA in water solution at 77 K: (1) DNA without the probe; (2) DNA with the 6SG probe (1/3000); (3) DNA with the 6SG probe (1/600); (4) absorption spectrum of the 6SG probe.

into account. Figure 5 also shows that there is good overlapping between the fluorescence spectrum of DNA and the absorption spectrum of 6SG. The calculation of the excitation transfer Forster radius gives $R_0 = 28$ Å. Both these results permit us to assume the induction-resonance mechanism of the excitation transfer to 6SG from the DNA molecule under UV radiation of the DNA, although there is a possibility of energy migration due to the exchange-resonance mechanism. These observations are of significant interest because 6SG acts as an antitumor drug. Our experiments show that the chemicotherapeutical action of 6SG can be modified (intensified) by additional UV irradiation of cancer cells.

Conclusions

1. 6-Thioguanosine (6SGR) is a suitable probe to study low and high molecular systems.

2. Phosphorescence intensity curves for 6SG in the native DNA manifest peculiarities at 21, 64, 87, 140, 182, and 268 K.

3. Vitrification (devitrification) of the hydrate shell of the major groove of DNA takes place in the range from 180 to 215 K. Cryoprotectors—DMSO and PD— added to the aqueous DNA solution suppress the vitrification temperature of DNA hydrate shell.

4. There is an effective energy transfer to the probe in the 6SG containing DNA after UV radiation.

Acknowledgments

Bibliography

[1] E. A. Burstein, *Proper Luminescence of Protein* Moskow, VINITI Ser. Biophysika 117 (1977) (In Russian).

[2] G. E. Dobretzov, in *Luminiscentnyi Analiz v Mediko Biologocheskykh Issledovaniyakh* (RMI, Riga, 1983), p. 72 (in Russian).

[3] Yu. V. Rubin, N. P. Bazhulina, V. A. Bokovoy, R. G. Lee, and V. O. Chekhov, Tezisy Dokl. VI. Vsesoyuzn. Konf. Po Spectroskopii Biopolimerov. Kharkov, 260 (1988) (in Russian).

[4] Yu. V. Rubin, F. A. Savin, and E. A. Leibina, in *Luminiscentnyi Analiz v MedikoBiologicheskikh Issledovaniyakh,* V. Sominsky, Ed. (RMI, Riga, 1990), p. 135 (in Russian).

[5] V. A. Bokovoy, Yu. V. Morozov, N. P. Bazhulina, and V. O. Chekhov, J. Phys. Khem. **65**, 865 (1991) (in Russian).

[6] P. Langen, *Antimetabolites of Nucleic Acid Metabolism.* (Gordon and Breach, New York, 1975).

[7] H. Kaplan, K. Smith, and P. Tomlin, Nature **2119**, 181 (1961).

[8] G. Lanselot, Mol. Phys. **31**, 241 (1976).

[9] P. Boutron and A. Kaufman, Cryobiology **16**, 557 (1979).

[10] G. Parks and H. Huffman, J. Phys. Chem. **31**, 1842 (1927).

[11] J. Ahlberg, E. Blanchard, and W. Lundberg, J. Chem. Soc. 539 (1937).

[12] O. Haida, H. Juga, J. Seki, Proc. Jpn. Acad. Sci. **48**, 683 (1972).

[13] E. F. Scheka, Usp. Fiz. Nauke **160**, 263 (1990) (in Russian).

[14] D. H. Rasmussen and A. Mackinze, Nature **220**, 1315 (1968).

[15] H. Drew, S. Samson, R. E. Dickerson, Proc. Natl. Acad. Sci. U.S.A. **79**, 4040 (1982).

[16] M. Kopka, A. V. Fratini, H. R. Drew, and R. E. Dickerson, J. Mol. Biol. **163**, 129 (1983).

[17] V. Zenger, *Printzypy Strukturnoj Organizatzii Nukleinovykh Kislot* (Mir, Moskow, 1987), p. 584 (in Russian).

[18] G. M. Mrevlishvili, Biophyzika **22**, 180 (1977) (in Russian).

[19] E. L. Andonikashvili, G. M. Mrevlishvily, G. S. Dzaparidze, and D. A. Tatishvily, Biophyzika **33**, 233 (1988) (in Russian).

Charged Particle Track Structure Parameters for Application in Radiation Biology and Radiation Chemistry

DAVID E. WATT AND ALI S. ALKHARAM

Radiation Biophysics, Department of Physics and Astronomy, University of St. Andrews, St. Andrews, Fife KY16 9LZ, Scotland, United Kingdom

Abstract

Chemical and biological damage, caused by directly or indirectly ionizing radiations, is attributable to the action of the charged particle tracks in the absorbing medium. Attempts to elucidate the biophysical mechanisms involved, and to quantify the damage, are typically made in terms of one or more of the main physical parameters descriptive of the charged particle tracks. To meet a need for a ready reference source of such information, tables of the relevant parameters have been calculated for a liquid water medium. The full tables are obtainable elsewhere. Here, a description is given of the quantities calculated and an extended example is given of their application in elucidating the physical mechanisms of radiation-induced biological damage. A representative selection of data is displayed graphically to illustrate the extent of the information obtained and its value in, e.g., application to fundamental radiation dosimetry. Track structure data is tabulated for instantaneous energies of individual particles and for the fluence and dose-weighted spectra at charged particle equilibrium. Data are listed for incident electrons (50 eV to 30 MeV); characteristic K_α X-rays from carbon to uranium; commonly used radioisotope sources of ^{241}Am, ^{137}Cs, and ^{60}Co and for continuous X-ray spectra ($\leq$300 kV); Auger electron and beta-emitter radionuclides; heavy charged particles having specific energies of 0.5 keV/μ to 1 GeV/μ for 74 ion types ranging from protons to uranium ions, and for monoenergetic neutrons (0.5 keV to 100 MeV). Quantities listed are kerma factors; fluence of charged particles per unit source concentration; buildup factors; track and dose-average LET and restricted LET; W values; z^2/β^2; β^2; delta-ray yields, energies, and ranges; ion ranges; and the mean free path for primary ionization and the linear primary ionization. For indirectly ionizing radiations, the microdose quantities, frequency, and dose means of lineal energy are tabulated along with typical energy deposition distribution spectra for neutrons and gamma rays in micron and nanometer volumes. © 1994 John Wiley & Sons, Inc.

1. Introduction

There is a continuing need for a comprehensive set of reference tables containing parameters descriptive of properties of the charged particle tracks generated directly or indirectly by ionizing radiations in biological materials. Applications are in diverse fields such as interpretation of biological damage mechanisms for radiological protection; modeling of radiation effects; heavy charged particle therapy; radiation dosimetry for astronauts and for computer components in space vehicles; assessment of effects of incorporated radionuclides in nuclear medicine; design of instrumental response for quality and dose specification; study and interpretation of inhomogeneous reaction-rate kinetics of induced radicals in aqueous media; and study of

damage induced in organic macro-molecules. Here, there is not space to reproduce the extensive sets of tables [1–3] but a brief description is given of the main parameters tabulated and of their physical meaning. Examples of the results for some of the quantities are displayed graphically. The results, calculated for a liquid water medium (representative of biological material), allow indirectly for physical phase and chemical-bonding effects.

A practical example is given of the use of the parameters in identifying fundamental mechanisms involved in the biological effectiveness of heavy charged particles. The tables entitled "Track structure data for ionizing radiations in liquid water" are in three main parts: Part 1 is on electrons and photons (50 eV to 30 MeV); Part 2 is on 74 types of heavy charged particles (100 eV/μ to 1 GeV/μ) ranging from protons to uranium ions; and Part 3 is on neutrons (0.5 keV to 100 MeV). The complete tables are available from the authors on request [1–3].

2. Details of the Calculated Track Structure Parameters and Their Physical Interpretation

2.1. Part 1(a): Electrons at Energies 50 eV to 30 MeV

Parameters calculated in the continuous slowing down approximation (CSDA) are the collisional linear energy transfer, LET (L_∞); the track and dose-average restricted LET (L_{100}); the linear primary ionization (I) and the mean free path (λ) between primary ionizations; the CSDA range (R); and the kerma factor (K_f). Results are given for electrons at instantaneous energies (e.g., for track segment experiments) and for electrons in charged particle equilibrium. The CSDA range of primary electrons has a dual meaning as it is also equivalent to the primary electron fluence per unit source concentration of electrons. Also given are the mean energies and ranges of the generation of secondary electrons in the equilibrium spectrum in addition to the parameters analogous to those given for the primary electrons and the relative variances of the dose and track LETs. The relative variance quantifies the width of the distribution.

Parameters listed that could have special significance in radiation chemistry are the space density of primary electrons in the initial track; the total space density of all electrons per unit source concentration of initial electrons; and the total electron fluence and factors for the buildup of electron concentration and of electron fluence. Approximate values are listed for the mean energy (W) expended in producing a primary ionization along primary electron tracks. Figure 1 shows the CSDA ranges for electrons in liquid water.

2.2 Part 1(b): Photons

Data, similar to those listed above for electrons, are tabulated for

(i) Characteristic (fluorescent) X-ray energies for 27 elements ranging from carbon to uranium. Figure 2 shows the relevant track-average LET, the dose-average LET, and the mean free path, λ, for primary ionization, for the equilibrium electron spectra generated by fluorescent K_α X-rays.
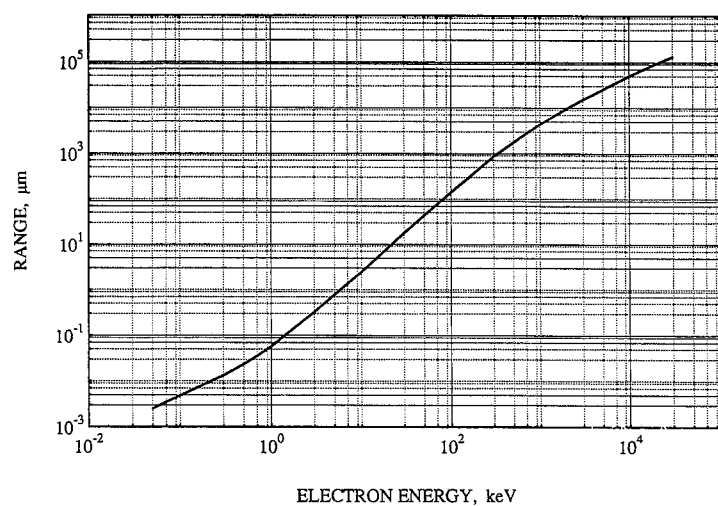
Figure 1. CSDA ranges for electrons in liquid water. Note that each decade on the log scale is "ticked" at 3, 5, 7, 9 on the grid.
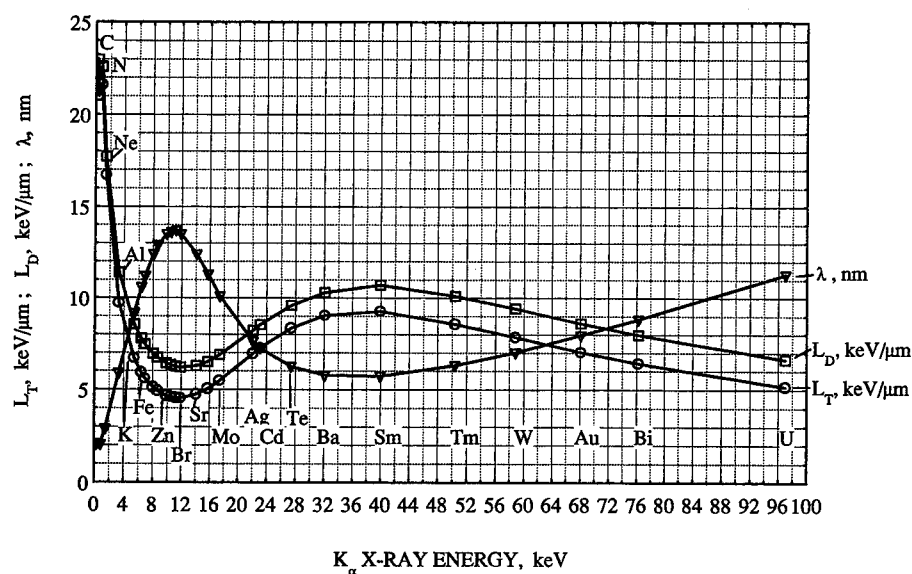

Figure 2. Track and dose-weighted LET and the mean free path for primary ionization, averaged over the equilibrium electron spectrum generated by characteristic X-rays in liquid water, are shown as a function of X-ray energy.

(ii) Orthovoltage continuous (medical) X-ray energy spectra for nine applied potentials ranging from 50 to 300 kV. Frequency-weighted and kerma-weighted mean photon energies, the mass energy transfer coefficients, the electron production cross sections, and the mean energies of photoelectrons and of Compton electrons are listed. The mass energy transfer coefficient combined with the initial photon energy, and the electron production cross sections combined with the electron spectrum, enable the validity of the calculation to be tested, as each should yield the same kerma factor (conservation of energy). This is proved to within 5%.

(iii) Radioisotope sources. Results are presented in the tables for Am-241 (59.6 keV γ), Cs-137 (661 keV γ), and Co-60 (1.17 and 1.33 MeV γ's).

(iv) Beta, Auger electron, and X-ray emitting radionuclides, viz., I-125, I-131, H-3, C-14, P-32, P-33, and Br-77. These radionuclides have wide application as tracers, as incorporated radionuclides, and in labeling of organic macromolecules. There is evidence that Auger-electron emitters are especially damaging.

The results are accurate to within 5% at electron energies above 10 keV but the accuracy progressively decreases to $\sim$ 10% at 1 keV energy and may be as large as 25% at 100 eV.

2.3. Part 2: Heavy Charged Particles/Accelerated Ions

Parameters for heavy charged particles, having specific energies between 1 keV per unit mass number and 1 GeV per unit mass number, have been calculated for 74 ion types ranging from protons to uranium ions. The ion types selected are those typically used in biological and chemical experiments with accelerated ions. Also included are light ions (lithium to oxygen isotopes) produced as typical fragmentation products in targets of low atomic number when bombarded by relativistic protons. The data have important application in space dosimetry and in high LET therapy. The calculations are based on Berger's recently revised results for stopping of protons and alpha particles (ICRU Report No. 49) [4] in liquid water. For fast ions, the track structure quantities calculated and tabulated as a function of specific energy have their origin in the Bohr/Bethe theories of stopping power (e.g., ICRU-37) [5], viz.:

$$\frac{1}{\rho} \cdot L_\infty \propto \frac{z^2}{\beta_i^2} \cdot F(\beta_i) \ .$$

β_i, the dimensionless ion velocity in atomic units, determines the maximum spatial distribution of delta rays around the ion track. z^2/β_i^2, the ratio of the effective charge to velocity, can be interpreted either as the linear yield of primary ionizations and/ or the linear yield of delta rays along the track. I, the linear primary ionization is the zeroth moment of energy transfer. Its reciprocal is the mean free path, λ, for primary ionization. L_∞, the collisional linear stopping power and the first moment of energy transfer, includes the kinetic energy of the radially distributed delta rays. The delta-ray properties depend only on the specific energy of the ions and are independent of ion type. A separate table on delta rays contains average and max-
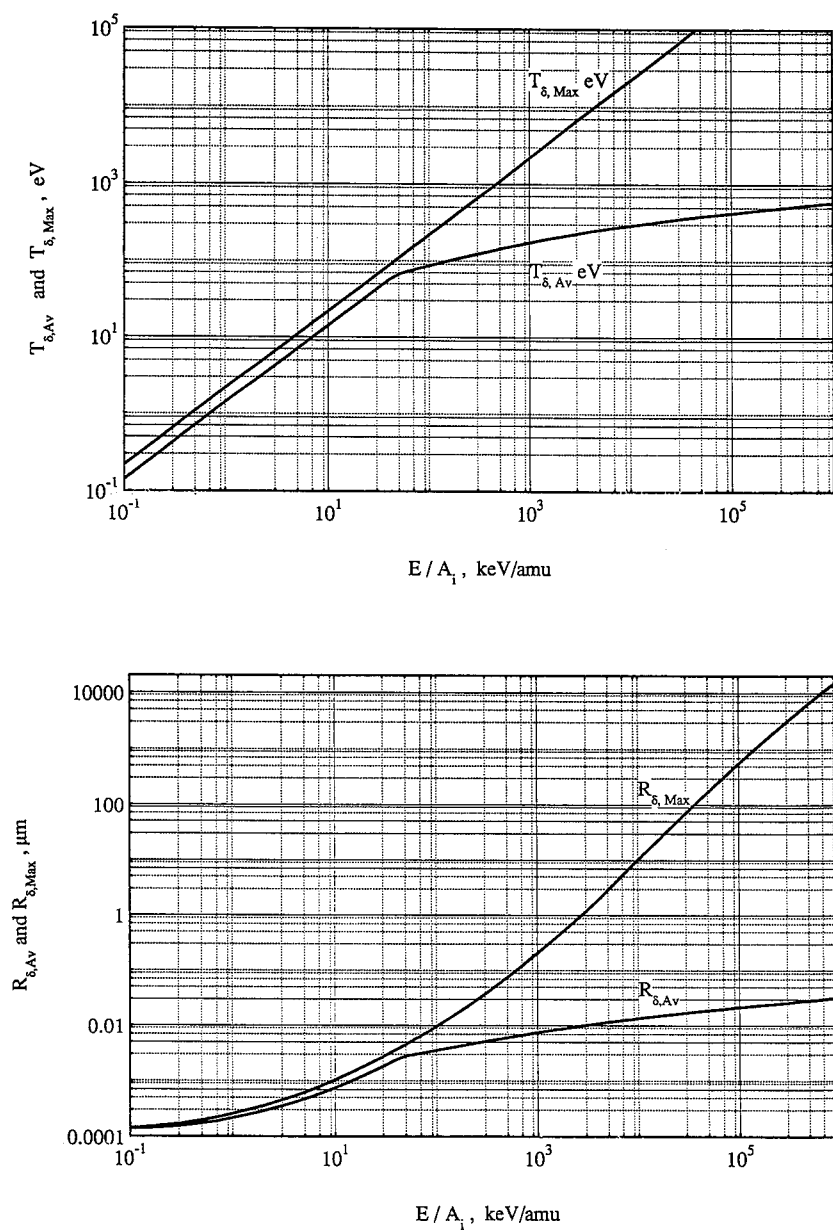
Figure 3.   (a) Maximum and average electron energies and (b) maximum and average ranges are shown for delta rays from heavy particle tracks. The limiting range for interpenetration between neighboring mammalian cells is $\approx$ 15 μm. Consequently, on average, only the portion of delta rays having energy $\geq$ 20 keV can enhance the effect cross section.

imum delta-ray energies and their corresponding CSDA ranges. These are illustrated
in Figure 3. Ion ranges are shown in Figure 4. W values for primary ionization are
tabulated. Analysis of the radiation effect for a specified end-point and ion type, in
terms of the three quantities β_i^2, z^2/β_i^2, and L_∞, should enable the components of
the total damage to be linked to the causal physical action, i.e., respectively, the
effects attributable to the spatial spread and number of delta rays, the spacing of
events, and the components of energy transfer. Restricted LET with 100 eV delta-
ray energy cutoff is also of interest in this context as it provides a measure of the
energy deposition along a cylindrical "core" of about 4 nm radius along the ion
track. Examples of the trends of the parameters are given in Figures 5–7. The
remaining energy loss is in the delta-ray penumbra. In radiation dosimetry, similar
quantities are required—but averaged over the charged particle equilibrium spec-
trum. Thus, the tables also contain the quantities track-averaged and dose-averaged
LETs and their restricted versions. The data tabulated for instantaneous specific
energies of ion are applicable to track segment experiments. Values for the equi-
librium spectrum of ions are applicable to irradiations where the whole track stops
in the biological material or where charged particle equilibrium may be achieved
as, e.g., in fast neutron irradiations.

2.4. Part 3: Neutrons (0.5 keV to 100 MeV)

In calculating the neutron data, only elastic collisions are considered. No allow-
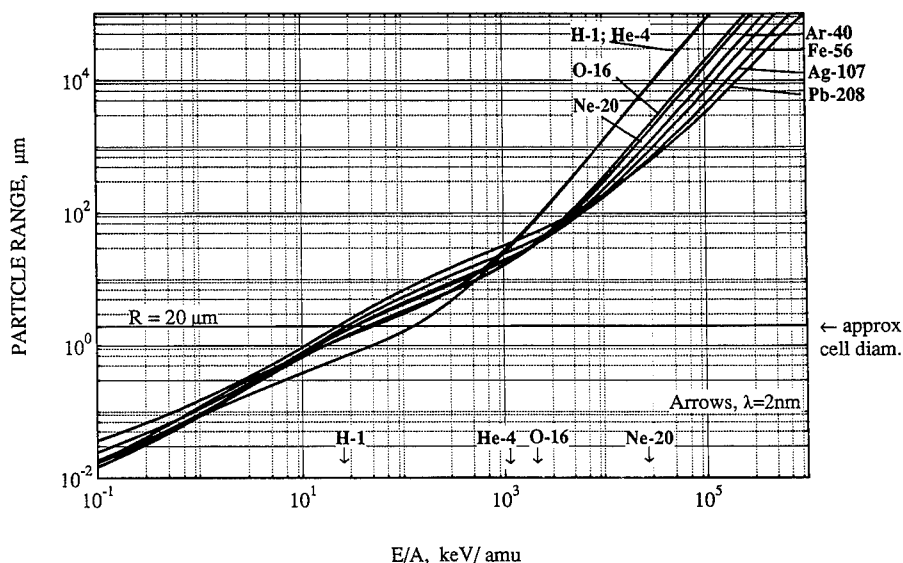ance is made for (n, α) reactions in oxygen (neutron threshold energy $\sim$ 6.3 MeV)

Figure 4.   Ranges of accelerated ions in liquid water. The range of an ion must be $\approx$ 20
μm to ensure track traversal of a mammalian cell. The vertical arrows along the abscissa
indicate specific energies corresponding to an ionization mean free path of 2 nm—at which
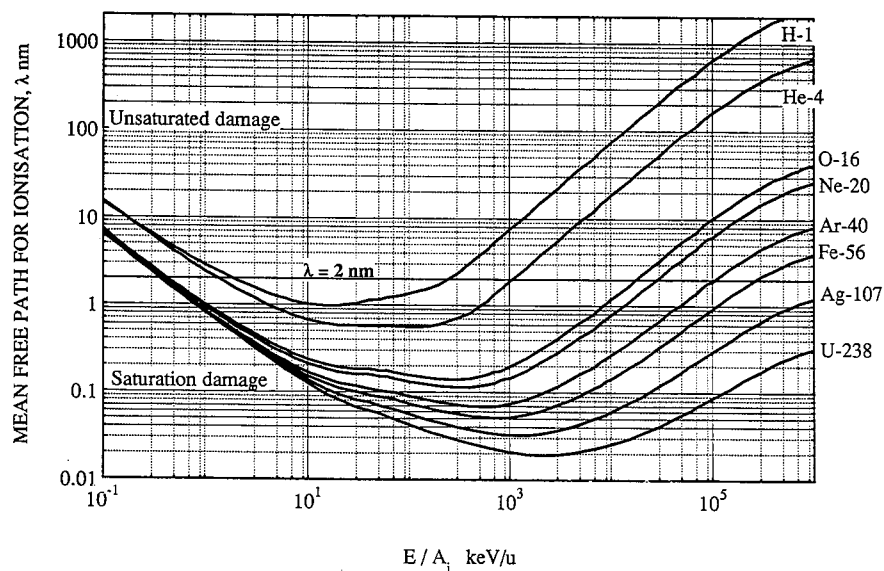RBEs will be maximum.

Figure 5. Mean free path for ionization is shown as a function of specific energy for the ion types indicated. For the study of damage mechanisms in biophysics, only light-accelerated ions (A ≤ 20) are appropriate; otherwise, basic mechanisms of interest will be obscured in the large background effects due to saturation damage.
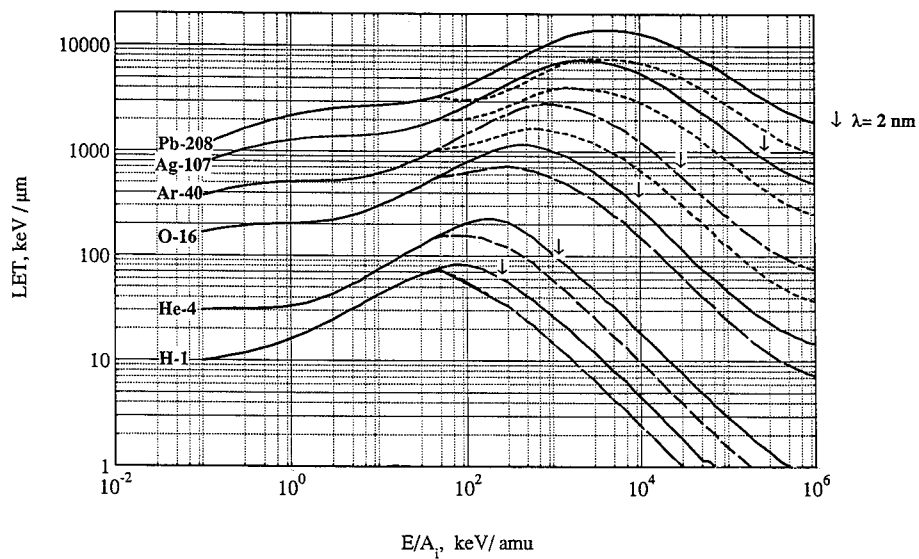

Figure 6. LET and restricted LET, L_{100}, for accelerated ions in liquid water.
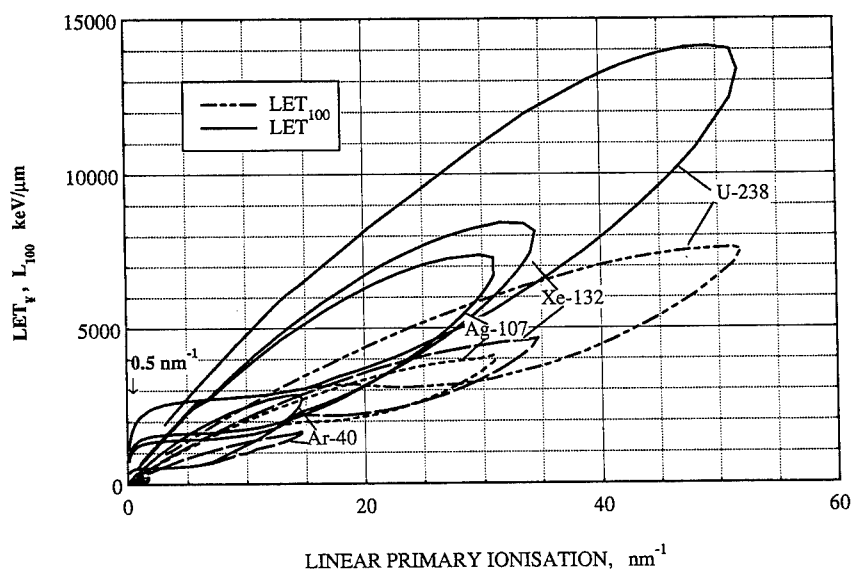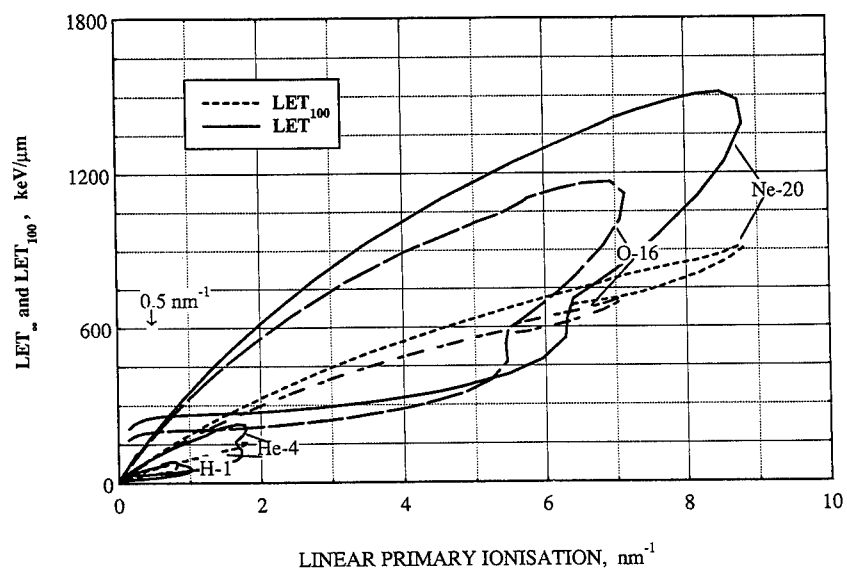
Figure 7. Shows the interrelationship among the three important quality parameters L_∞, L_{100}, and $I(\propto z^2/\beta^2$ for fast ions). Note that for biophysical modeling nearly all of the ion types are in the unwanted saturation region ($I > 0.5$ nm) for mammalian cells. (a) Low Z ions; (b) high Z ions.

or for multiparticle breakup reactions (neutron threshold energy ~ 20 MeV). For radiological protection, the neutron energies most commonly experienced are in the keV to 15 MeV energy range. Biophysical data calculated for both the H and O recoil component produced by neutron interaction in water are the recoil source density, the partial kerma factors, and the equilibrium fluence of recoils of H and O per unit incident neutron fluence. Track structure quantities listed are those indicated for the other radiations described above plus maximum and average recoil energies, their projected ranges, the maximum and average delta-ray energies, and ranges associated with the recoil tracks.

In recent years, the subject of microdosimetry, which takes into account the stochastic nature of energy deposition, has been applied in attempts to interpret radiation effects and mechanisms in small sites [39]. Site sizes were typically of micron dimensions to represent the dimensions of biological cell nuclei and of inhomogeneous rate processes in biochemistry. More recently, the possible importance of nanometer sites in macromolecular targets has been recognized, e.g., in the DNA. Relevant information on the frequency and dose distributions and on the corresponding mean lineal energies along with the relative variances is provided. Figure 8 shows the track and dose-weighted mean lineal energies in water for the H recoil component generated by monoenergetic neutrons in a 1 μm site diameter.

3. Application of Track Structure Parameters in the Interpretation of Damage to Mammalian Cells

Damage by ionizing irradiations of mammalian cells is commonly expressed in terms of the surviving fraction as a function of absorbed dose for a specified biological
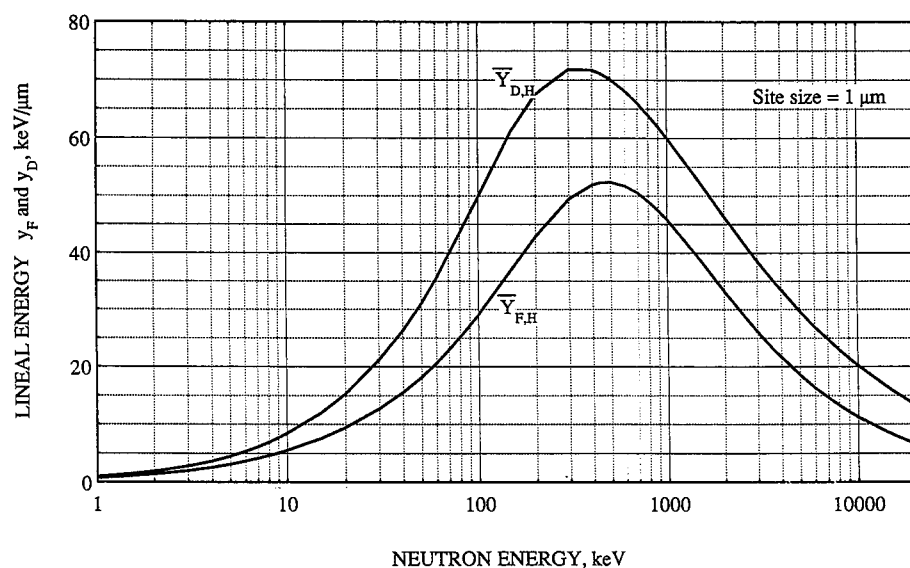

Figure 8.   Frequency and dose-weighted lineal energies, y_F keV/μm and y_D keV/μm, are shown for monoenergetic neutrons in liquid water.
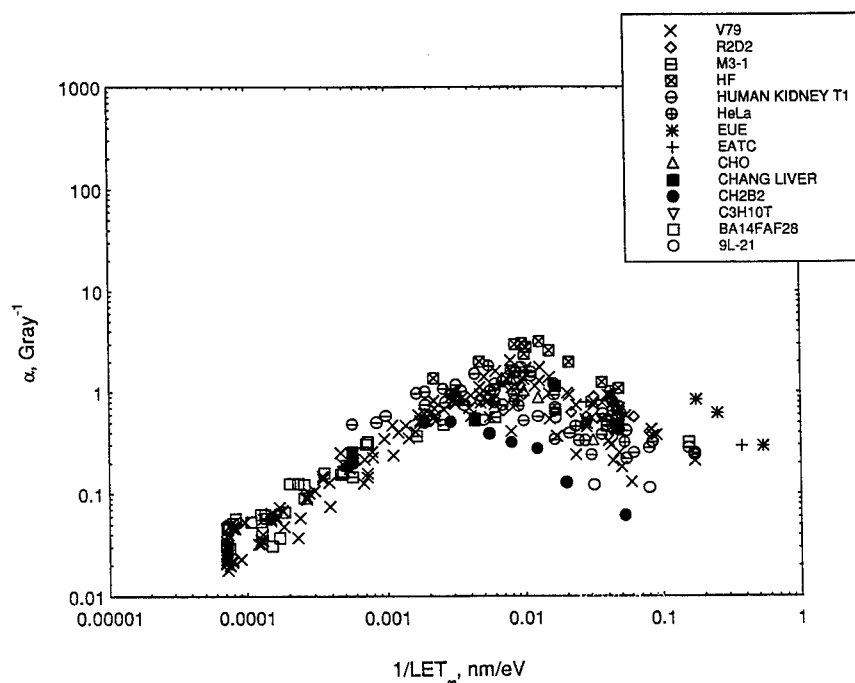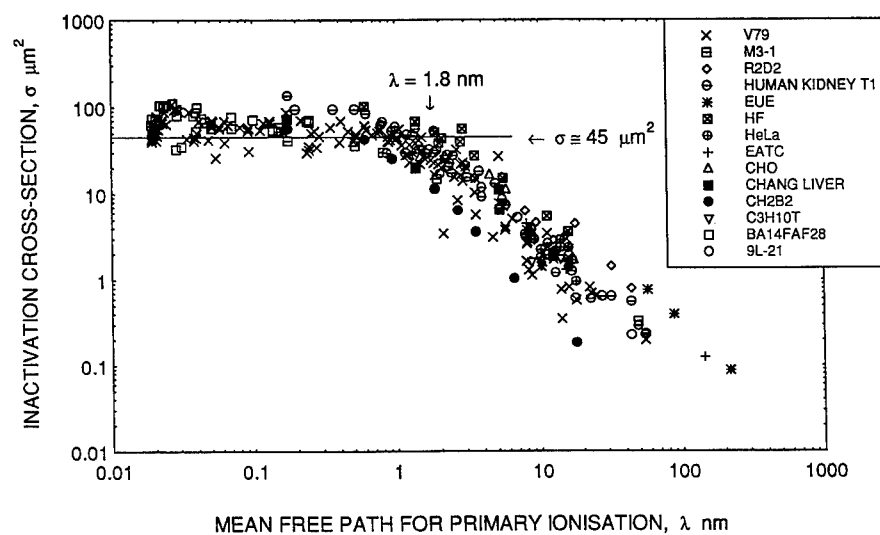
Figure 9.   Interpretation of damage mechanisms for inactivation of mammalian cells. (a) is shown to be more meaningful and the damage coefficients better correlated as a function of ionization mean free path, λ nm, than (b) as a function of L_∞. Data are extracted from [6-38].

end-point. The information obtained is of value for interpretation of damage mechanisms implicit in the development of models of radiation action used in radiological protection. Good models are required to permit reliable extrapolation of effects from the high dose levels used in laboratories to the lower doses and dose rates occurring near environmental levels and experienced by the majority of the population. Parameters can be identified that are of value in quantifying, e.g., the risk of radiation-induced cancer. Figure 9(a) shows a compilation of results, extracted from the literature, for the initial slope "α" in gray^{-1} as a function of mean LET, L keV/μm, for mammalian cell survival after irradiation with a wide selection of heavy ion types. This graph can be considered as a representation of the currently accepted system of dosimetry used to determine exposure limits in safety legislation. Close examination of the plotted curve shows that the saturation value of α occurs at different values of L, depending on charged particle type, and, consequently, there is no possibility of obtaining a unified curve with these parameters. The currently accepted system of dosimetry is therefore fundamentally unsound. Since the cross section for induction of the specified biological end-point represents the probability of the effect per unit fluence of tracks, a more appropriate procedure is to compare the effect cross section with the track parameters of interest. If a "unified" plot can be obtained for a wide range of radiation types, then the relevant parameter must surely be of significance. The effect cross section, σ_{eff}, is related to the initial slope of the dose-survival curve by

$$\sigma_{eff}(\text{cm}^2) = \frac{1.6 \times 10^{-9} \cdot \bar{L}_T(\text{keV}/\mu\text{m})}{\rho(\text{g/cm}^3) \cdot D(\text{Gy})},$$

where L_T is the track average LET; D, the absorbed dose, and ρ, the density. For the analysis, the initial slope is selected to avoid any complications due to subsequent recovery of damage that may be associated with cellular repair processes. The data used in Figure 9(a) are transposed in Figure 9(b) to a $\sigma_{eff} - \lambda$ plot, using the equation above. Here, λ is the mean free path for *primary* ionization along the charged particle tracks. Several interesting factors are revealed. There is a significantly better correlation of results onto a unified curve. There is a common point of inflexion, for all charged particle types, at a mean free path of 2 nm, which, on various arguments, leads to the conclusion that the key lesion is damage to the DNA. The spacing of events, rather than energy deposition, is the mechanism involved. As the σ–λ plot is better correlated than is the α–L plot, the conclusion is that delta rays (the energy of which are included in the LET) must have much reduced, probably negligible, effect except in the highly saturated region. This finding challenges the physical basis of the system of "dose-limitation" used in radiation protection and the assumption that energy deposition is the important quantity in determining damage. Numbers and the correlated spacing of pairs of interactions, whether direct or indirect, are found to be the key factors that are common to all radiation types.

Acknowledgments

Bibliography

[1] D. E. Watt, *Track structure data for ionizing radiations in liquid water, Part 1: electrons and photons.* St. Andrews University Report Biophys/10/89 (1989).

[2] D. E. Watt, *Track Structure Data for Ionizing Radiations in Liquid Water, Part 2: Heavy Charged Particles 100 eV/A to 1 GeV/A.* St. Andrews University *Report* Biophys/1/93 (1993).

[3] D. E. Watt, *Track Structure Data for Ionizing Radiations in Liquid Water, Part 3: Neutrons 0.5 keV to 100 MeV.* St. Andrews University Report Biophys/11/93 (1993).

[4] *Stopping Power and Ranges for Protons and Alpha Particles.* ICRU Report No. 49. International Commission in Radiation Units and Measurements, 7910 Woodmont Ave., Bethesda, MD 20814 (1993).

[5] *Stopping Power for Electrons and Positrons.* ICRU Report No. 37. International Commission in Radiation Units and Measurements, 7910 Woodmont Ave., Bethesda, MD 20814 (1984).

[6] G. W. Barendsen, Int. J. Radiat. Biol. **8**(5), 453 (1964).

[7] G. W. Barendsen, C. J. Koot, G. R. van Kersen, D. K. Bewley, S. B. Field, and C. J. Parnell, Int. J. Radiat. Biol. **10**(4), 317 (1966).

[8] G. W. Barendsen, H. M. D. Walter, J. F. Fowler, and D. K. Bewley, Radiat. Res. **18**, 106 (1963).

[9] M. Belli, F. Gera, R. Cherubini, F. Ianzini, G. Moschini, O. Sapora, G. Simone, M. A. Tabocchini, and P. Tiveron, Int. J. Radiat. Biol. **61**(1), 145 (1992).

[10] M. Belli, R. Cherubini, S. Finotto, G. Moschini, O. Sapora, G. Simone, and M. A. Tabocchini, Int. J. Radiat. Biol. **55**(1), 93 (1989).

[11] U. Bertsche, and G. Iliakis, in *7th Symposium on Microdosimetry EUR 7147,* CEC, Brussels (1981), pp. 1251–1262.

[12] D. Bettega, C. Birattari, M. Bombana, A. M. Fuhrman Cont, E. Gallini, T. Pelucchi, and L. T. Lombardi, Radiat. Res. **77**, 85 (1979).

[13] R. P. Bird, and H. J. Burki, Int. J. Radiat. Biol. **27**(2), 105 (1975).

[14] E. A. Blakely, C. A. Tobias, T. C. H. Yang, K. C. Smith, and J. T. Lyman, Radiat. Res. **80**, 122 (1979).

[15] P. Chang, C. A. Tobias, and E. A. Blakely, Radiat. Res. **129**, 272 (1992).

[16] R. Cox, J. Thacker, D. T. Goodhead, and R. J. Munson, Nature **267**(2), 425 (1977).

[17] R. Cox and W. K. Masson, Int. J. Radiat. Biol. **36**(2), 149 (1979).

[18] R. A. Deering and R. Rice, Radiat. Res. **17**, 774 (1962).

[19] M. Folkard, K. M. Prise, B. Vojnovic, S. Davies, M. J. Roper, and B. D. Michael, Int. J. Radiat. Biol. **56**(3), 221 (1989).

[20] D. T. Goodhead, M. Belli, D. A. Bance, L. A. Allen, S. C. Hall, F. Ianzini, G. Simone, D. L. Stevens, A. Stretch, M. A. Tabocchini, and R. E. Wilkinson, Int. J. Radiat. Biol. **61**(5), 611 (1992).

[21] G. Kraft, W. Kraft-Weyrather, H. Meister, H. G. Miltenburger, R. Roots, and H. Wulf, in *8th Symposium on Microdosimetry, EUR 8395,* CEC, Brussels (1983), pp. 743–753.

[22] T. Kranert, E. Schneider, and J. Kiefer, Int. J. Radiat. Biol. **58**(6), 975 (1990).

[23] T. Kranert, E. Schneider, and J. Kiefer, in *Proceedings of the 21st Annual Meeting of the European Society of Radiation Biology* (1988).

[24] T. Kranert, U. Stoll, E. Schneider, and J. Kiefer. Adv. Space Res. **12**(2,3), 111 (1992).

[25] R. J. Munson, D. A. Bance, A. Stretch, and D. T. Goodhead, Int. J. Radiat. Biol. **36**(2), 127 (1979).

[26] F. Q. H. Ngo, E. A. Blakely, and C. A. Tobias, Radiat. Res. **87**, 59 (1981).

[27] C. F. Perez, M. R. Botchan, and C. A. Tobias, Radiat. Res. **104**, 200 (1985).

[28] A. Perris, P. Pialoglou, A. A. Katsanos, and E. G. Sideris, Int. J. Radiat. Biol. **50**(6), 1093 (1986).

[29] M. R. Raju, Y. Eisen, S. Carpenter, and W. C. Inkret, Radiat. Res. **128**, 204 (1991).

[30] C. J. Roberts, and D. T. Goodhead. Int. J. Radiat. Biol. **52**(6), 871 (1987).

[31] A. Rodrigues and E. L. Alpen. Radiat. Res. **85**, 24 (1981).

[32] L. D. Skarsgard, B. A. Kihlman, L. Parker, C. M. Pujara, and S. Richardson, Radiat. Res. Suppl. **7**, 208 (1967).

[33] J. Thacker, A. Stretch, and D. T. Goodhead, Radiat. Res. **92**, 343 (1982).

[34] J. Thacker, A. Stretch, and M. A. Stephens, Int. J. Radiat. Biol. **36**(2), 137 (1979).

[35] P. Todd, Radiat. Res. **34**, 378 (1968).

[36] P. Todd, Med. Coll. Virginia Q. I, 2 (1965).

[37] P. Todd, Radiat. Res. 61, 288 (1975).

[38] H. Wulf, W. Kraft-Werather, H. G. Miltenburger, E. A. Blakely, C. A. Tobias, and G. Kraft, Radiat. Res. 104, S-122 (1985).

[39] Microdosimetry. ICRU Report No. 36. International Commission in Radiation Units and Measurements, 7910 Woodmont Ave., Bethesda, MD 20814 (1983).

Correlation of the Partition Coefficient with the Molecular Transform Index in Series of Organophosphorus Compounds

JAMES W. KING

US Army Edgewood Research, Development and Engineering Center,
Aberdeen Proving Ground, Maryland 21010-5423

Abstract

The molecular transform index (FT_m) is a unitary representation of a molecule based on a Fourier operation on the bond distance or graphical descriptor matrices of the structure while incorporating the atomic number of the constituent atoms. In a series consisting of phosphonates, phosphonothionates, and a phosphinate, the FT_m gave an excellent linear correlation ($R = 0.91$) with experimentally determined octanol/water partition coefficients (log $P_{o/w}$). In a second group containing phosphonofluoridates, thionophosphonofluoridates, and phosphoramidofluoridates, the FT_m correlation with log $P_{o/w}$, calculated by the π-fragment method, served to separate the fluoridates and amidates as structural subclasses. © 1994 John Wiley & Sons, Inc.*

Introduction

The molecular transform index (FT_m) is a unitary numerical representation of a molecular structure. It is determined in four steps by performing a Fourier operation on any one of several possible structure representations to give a curve analogous to that of free induction decay, squaring the points on the curve to make all the amplitudes positive, integrating the area under the curve, and then taking the square root of the area. The structure may be represented by a matrix of Cartesian coordinates or virtually any graphical depictions such as adjacency or distance matrices, and the latter may be either distances between bonded atoms only (two-dimensional) or between all atoms (three-dimensional). In essence then, the generation of the index is a mapping-down process from a two- or three-dimensional to a one-dimensional space. However, the structure is even more precisely defined by including the atomic number of the constituent atoms in the original Fourier operation [1]. In several studies, the index has been used to correlate both physical and pharmacological properties in a series of highly varied structures [1]; two thermodynamic functions in a series of linear and branched hydrocarbons [2]; bivariant enzyme inhibition activity in a large series of organophosphorus compounds [3];

and, in an alternate form, as a descriptor for molecular similarity analyses [4]. The FT_m compares very favorably with descriptors derived from quantum chemical (MNDO) calculations [5].

The successful use of the FT_m for the correlation of a variety of indices led naturally to considering its value as a correlator, or even predictor, of the "extra-thermodynamic" molecular parameter, the partition coefficient. This parameter, reflecting hydrophobicity [6], is highly indicative of membrane as well as general transport potential in biological systems and in its functional form is expressed as the logarithm$_{10}$ (log P) of the octanol/water partition coefficient as measured under rather standard conditions [7–9]. Hansch and Leo developed an empirical paradigm for its calculation known as the "π-fragment" method [10]. While there have been variations on this methodology, perhaps the most promising from a basic theoretical standpoint is that of Politzer and Murray, in which the calculation is based on molecular surface area and electrostatic potentials [11–13]. However, inasmuch as the FT_m is also structure-based, and rather less complex to calculate, its value as a correlator for log P was considered to be appropriate for investigation. For this purpose a series of organophosphorus compounds was selected whose partition coefficients had been measured by High Performance Liquid Chromatographic methodology (Table I) [14]. For an evaluation of FT_m correlation capability *versus* partition coefficients calculated by the π-fragment method, the compounds shown in Table II were chosen.

Methodology and Results

The log P values shown in Table II were calculated by the π-fragment methodology as described for examples cited by Leo, Hansch, and Elkins [15]. The fragment

TABLE I.   HPLC-measured partition coefficients (log P) and FT_m values.

Compound No.	Structure	Log P^a	$FT_m{}^b$ ($\times 10^2$)
1	$CH_3P(O)(H)OC_2H_5$	−0.60	2.021
2	$CH_3P(O)(OCH_3)_2$	−0.61	2.535
3	$C_2H_5P(O)(OC_2H_5)_2$	0.66	3.208
4	$CH_3P(O)(OC_2H_5)(SC_2H_5)$	0.71	3.226
5	$CH_3P(O)[OCH(CH_3)_2]_2$	1.03	3.435
6	$CH_3P(S)(OC_2H_5)_2$	2.08	3.120
7	$CH_3P(O)(OC_4H_9)_2$	3.26	3.891
8	$CH_3P(S)(OC_3H_7)_2$	3.26	3.074
9	$CH_3P(O)(OC_5H_{11})_2$	3.74	4.350
10	$CH_3P(O)[OCH(CH_3)C(CH_3)_3]_2$	4.29	4.811
11	$CH_3P(O)(OC_8H_{17})_2$	6.13	5.735

[a] Taken from Ref. [14].
[b] Molecular transform index.

TABLE II. Calculated partition coefficients (Clog P) and FT_m values.

Compound No.	Structure	CLog P[a]	FT_m[b] ($\times 10^2$)
1	$CH_3P(O)(F)O$-iso-propyl	0.54	2.801
2	$CH_3P(O)(F)OCH(CH_3)C(CH_3)_3$	1.84	3.488
3	$CH_3P(O)(F)O$-cyclohexyl	2.04	3.596
4	$CH_3P(O)(F)O$-(2-methylcyclohexyl)	2.34	3.826
5	$CH_3P(S)(F)OCH(CH_3)C(CH_3)_3$	2.30	3.442
6	$CH_3P(S)(F)O$-(2-methylcyclohexyl)	2.50	3.908
7	$CH_3P(S)(F)O$-cyclopentyl	2.00	3.306
8	$CH_3P(O)(F)O(CH_2)_2N(iso$-propyl$)_2$	1.54	4.269
9	$CH_3P(O)(OCH_2CH_3)SCH_2CH_2N(CH(CH_3)_2)_3$	1.57	4.830
10	$(CH_3O)_2P(O)$-(1-morpholinyl)	−2.05	4.006
11	$(CH_3)_2NP(O)(F)O(CH_2)_2N(CH_3)_2$	−0.86	3.970
12	$(CH_3)_2NP(O)(F)O(CH_2)_2N(iso$-propyl$)_2$	0.64	4.876
13	$(CH_3)_2NP(O)(F)O(CH_2)_3N(CH_3)_2$	−0.36	4.157
14	$(CH_3)_2NP(O)(F)S(CH_2)_2N(iso$-propyl$)_2$	2.02	5.017
15	$(CH_3)_2NP(O)(F)(O(CH_2)_2$-(1-pyrrolidinyl))	−0.56	4.484
16	$(CH_3)_2NP(O)(F)(O$-3-(1-methyl-pyrrolidinyl))	−1.05	4.293
17	$(CH_3)_2NP(O)(F)(O$-(3-quinuclidinyl))	−0.05	4.888

[a] Calculated according to Reference [10].
[b] Molecular transform index.

values for the phosphoryl and analogous thiono entity were determined from the reported measured partition coefficients of paraoxon and parathion [10], respectively, by "backing out," i.e., subtracting, the appropriate π-fragments; it was then possible to "add in" the π-fragments necessary to construct the compound of interest and thereby arrive at its log P value.

The FT_m values were calculated from matrices of bond distances and atomic numbers of constituent atoms of the molecules as previously reported [1–4]; in these citations the FT_m was denoted as SQRT. The bond distances were taken from Gordon and Pople [16], March [17], and Corbridge [18]. The correlation trials were performed on a Texas Instruments TI-59 Programmable Calculator using a regression program described by Clark [19].

Correlation Analyses

Linear regression of the log P and FT_m values for the compounds shown in Table I gave the following:

$$\log P = 0.0186\,FT_m - 4.472 \tag{1}$$

$$n = 11 \quad R = 0.911 \quad S = 0.931 \quad F = 44.176$$

The compounds in Table II were considered in two groups as shown below:
Compounds 1–7:

$$\log P = 0.0163 FT_{\mathrm{m}} - 3.722 \tag{2}$$

$$n = 7 \quad R = 0.909 \quad S = 0.299 \quad F = 23.839$$

Compounds 11–16:

$$\log P = 0.0247 FT_{\mathrm{m}} - 11.068 \tag{3}$$

$$n = 6 \quad R = 0.872 \quad S = 0.636 \quad F = 12.745$$

where:

n = number of compounds,
R = correlation coefficient,
S = standard deviation,
F = F statistic.

For the compounds in the first group in Table II (nos. 1–7), inclusion of compound 8 in the regression dropped the correlation coefficient (R) to 0.543. Inclusion of compound no. 17 in the second group resulted in an R value of 0.800. Compounds 9 and 10 did not fit into either correlation grouping.

Discussion

The success of the molecular transform index as a unitary structure descriptor for the analysis and prediction of structural features, physicochemical and thermodynamic properties, and pharmacological activity, as shown in previous studies, serves to underline its viability in the present instance. Nowhere is this more evident than in the correlation for the compounds in Table I. In this case, the partition coefficients had been measured with a high degree of precision and this is reflected in the relatively high correlation coefficient even though the series spans several orders of magnitude of the partition coefficient itself. Contrarily, the correlations for the compounds shown in Table II, while acceptable for preliminary estimations, may be indicative of the empiricity or limitations of the particular partition coefficient calculation method. In respect to log P estimation it is known that, for polar molecules such as the organophosphorus compounds, parameter estimation is not well established and in this case was based upon only two actual measurements. From a structure correlation viewpoint the FT_{m} very nicely separated the fluoridates and the phosphoramides and, indeed, a rough plot of log P *versus* the FT_{m} values shows this separation as two approximately parallel classes.

The log P estimation problem for polar compounds is quite evident when compound 8 is included with compounds 1–7 of Table II as the correlation declines precipitiously. This degradation is less drastic when compound 17 is included with numbers 11 through 16. However, accounting for the lack of fit of compounds 9 and 10 in either regression is difficult. One may argue that 10 is a phosphate, the lone member of a class not elsewhere considered in this study. From the general impetus of the data herein, it would appear likely that the log P values for compound 9, and perhaps 8 also, were underestimated. In any case, further studies with these and other classes of compounds is warranted.

Conclusions

The molecular transform index (FT_m) has been shown to be a useful unitary structure descriptor for correlation of the octanol/water partition coefficient and structure in series of organophosphorus compounds. This was most evident where the log P values for a series containing phosphonates, phosphonothionates, and a phosphinate were experimentally measured with precision. The FT_m also served to functionally discriminate two series for which the log P was calculated by the Leo/Hansch π-fragment methodology. One contained phosphonofluoridates and thionophosphonofluoridates, and the other predominantly phosphoramides.

Acknowledgments

Bibliography

[1] J. W. King and R. J. Kassel, Int. J. Quantum Chem., Quantum Biol. Symp. **17**, 27 (1990).
[2] J. W. King and R. J. Kassel, Int. J. Quantum Chem., Quantum Biol. Symp. **18**, 289 (1991).
[3] J. W. King and R. J. Kassel, Int. J. Quantum Chem., Quantum Biol. Symp. **19**, 179 (1992).
[4] J. W. King, Int. J. Quantum Chem., Quantum Biol. Symp. **20**, 139 (1993).
[5] G. R. Famini, R. J. Kassel, J. W. King, and L. Y. Wilson, Quantum Struct.-Act. Relat. **10**, 344 (1991).
[6] A. J. Leo, Hydrophobicity, The Underlying Property in Most Biochemical Events, in *Environmental Health Chemistry*, J. D. Mckinney, Ed. (Ann Arbor Science Publishers, Inc., Ann Arbor, MI, 1981), pp. 323–336.
[7] W. P. Purcell, G. E. Bass, and J. M. Clayton, *Strategy of Drug Design: A Guide to Biological Activity* (Wiley, New York, 1973), pp. 126–143.
[8] S. H. Unger, P. S. Cheung, G. H. Chiang, and J. R. Cook, RP-HPLC Determination of 1-Octanol Partition and Distribution Coefficients: Experience and Results, in *Partition Coefficient Determination and Estimation*, W. J. Dunn, III, J. H. Block, and R. S. Pearlman, Eds. (Pergamon Press, New York, 1986), pp. 83–99.
[9] D. A. Brent, J. J. Sabatka, D. J. Minick, and D. W. Henry. J. Med. Chem. **26**, 1014 (1983).
[10] C. Hansch and A. J. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology* (Wiley, New York, 1979).
[11] P. Politzer, J. S. Murray, M. C. Concha, and T. Brinck, J. Mol. Struct. (Theochem) **281**, 107 (1993).
[12] J. S. Murray, T. Brinck, and P. Politzer, J. Phys. Chem. **97**, 13807 (1993).
[13] T. Brinck, J. S. Murray, and P. Politzer, J. Org. Chem. **58**, 7070 (1993).
[14] S. E. Krikorian, T. A. Chorn, and J. W. King, Quantum Struct.-Act. Relat. **6**, 65 (1987).

[15] A. J. Leo, C. Hansch, and D. Elkins, Chem. Rev. **17**(6), 525 (1971).
[16] M. S. Gordon and J. A. Pople, Quantum Chem. Program Exchange (QCPE) **11**, 135 (1969).
[17] J. March, *Advanced Organic Chemistry,* 4th ed. (Wiley, New York, 1992), Chap. 1.
[18] D. E. C. Corbridge, *Phosphorus: An Outline of Its Chemistry, Biochemistry, and Technology,* 4th ed. (Elsevier Science Publishers, Amsterdam, The Netherlands, 1990).
[19] F. H. Clark, *Calculator Programming for Chemistry and the Life Sciences* (Academic Press, New York, 1981), Chap. 4.

Curve-Fitting Paradox

MILAN RANDIĆ

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311

Abstract

We consider the paradoxical situation arising in the standard multiple regression analysis in that as the standard error of prediction decreases by introduction of additional variables (descriptors) at the same time the standard error of the coefficients of the regression analysis increases, often to the point of the coefficients having no statistical validity. We trace the origin of this paradoxical situation to inter-correlation of the variables. A remedy to this curve-fitting paradox is in the introduction of orthogonal variables or descriptors. © 1994 John Wiley & Sons, Inc.

Introduction

An apparent contradiction observed in the rigorous sciences is referred to as a paradox. A paradox may be a consequence of a lack of adequate conceptual basis, as is illustrated by the paradox of Zeno, one of the best-known paradoxes from antiquity. Zeno posed it as the problem of a race between the old Greek mythological athlete Achilles and the tortoise. Old Greeks had no notion of the limit to recognize that while there is no end to the sequence of hypothetical steps of Achilles to reach the current position of the tortoise, which was given an initial advantage, the constructed sequence has a limit.

Often, the term paradox is used in a less rigorous sense. Thus, the so-called Creten Paradox illustrates an apparent but not a genuine paradox. Philosopher Epimenides is credited to have said: "All Cretans are liars." Was he, being a Cretan, lying or speaking truth? The statement is not a true paradox since it may represent a false statement [1].

The apparent and genuine contradictions often have their origins in the meta-mathematical context of the statements that refer to themselves. In a way, the paradox that we will discuss in this article has elements of self-reference. We will focus on an apparent paradox of numerical content, involving statistical analysis. Consider curve fitting either by a power series or a closely related multiple linear regression analysis. In such problems, one considers a set of powers $1, f(x), f^2(x), f^3(x), \ldots f^k(x)$, the simplest cases being $f(x) = x$ or $f(x) = 1/x$, to represent a curve, or a set of descriptors $d_1, d_2, d_3, \ldots d_k$ to represent a plane in k-dimensional

---

This article is dedicated to a dedicated Swede, Professor Yngve Öhrn, who continues to enrich us with advances in mathematical physics and mathematical chemistry.

space. We can consider different powers x^k or $f^k(x)$ as different descriptors d_k to represent a property P of a molecule

$$P(G_n) = c_1 d_1 + c_2 d_2 + c_3 d_3 + \ldots + c_k d_k, \qquad (1)$$

which reduces the power expansion to multiple regression analysis. Here, G_n is the n-th object (molecule, graph) in a set of which data are known and to be fitted by d_k descriptors. The coefficients c_k are determined by the least-square procedure.

The precision of the fitting of the data is measured by the standard error S and the associated coefficient of regression R or its square R^2 (the coefficient of the determination). Statistically significant descriptors tend to decrease the standard error and increase the coefficient of the regression R. In the ideal situation, the limiting value for S is zero, while R approaches one.

All this is well known. However, an analysis of the statistical parameters of the regression equation, i.e., an analysis the coefficients c_k appearing in a regression equation, often was overlooked. When undertaken, it was observed with great disappointment that an increase in the number of descriptors (or the number of power terms, in the case of power series expansion) worsen the standard errors for the coefficients of a regression equation dramatically. That means that the equations are becoming less reliable as more descriptors were used. By the time one reaches a satisfactory standard error S for the property P, the standard errors for the corresponding coefficients in the equation of regression show that the coefficients are statistically meaningless. So, we have a paradoxical situation: To obtain reliable predictions for properties, the analysis rests on statistically "unreliable" equations.

Illustration

Let us illustrate the paradox on the regression equation derived for the 18 isomers of octane for their heats of formation (H_f):

$$H_f = 1.41 - 10.08\chi_1 - 4.86\chi_2 - 4.50\chi_3$$

$$- 0.04\chi_4 - 1.34\chi_5 + 2.01\chi_6 + 5.83\chi_7. \qquad (2)$$

Here, the descriptor χ_1 is the connectivity index [2] and the χ_k are the higher connectivity indices [3]. The connectivity indices are mathematical invariants that can be evaluated once the molecular structural formula is known. The connectivity index χ_1 is given as a sum of weighted CC bond contributions. The weighing takes into account the presence of the neighboring carbon atoms and is defined by the following numerical values [2]:

	CC bond weights	Numerically
Primary–secondary	$1/\sqrt{1 \cdot 2}$	0.70711
Primary–tertiary	$1/\sqrt{1 \cdot 3}$	0.57735
Primary–quaternary	$1/\sqrt{1 \cdot 4}$	0.50000
Secondary–secondary	$1/\sqrt{2 \cdot 2}$	0.50000

Secondary–tertiary	$1/\sqrt{2 \cdot 3}$	0.40825
Secondary–quaternary	$1/\sqrt{2 \cdot 4}$	0.35355
Tertiary–tertiary	$1/\sqrt{3 \cdot 3}$	0.33333
Tertiary–quaternary	$1/\sqrt{3 \cdot 4}$	0.28868
Quaternary–quaternary	$1/\sqrt{4 \cdot 4}$	0.25000

With the above definition of the bond contributions, one can easily compute the connectivity indices for molecules of interest. When so calculated, connectivity indices are taken as molecular descriptors in a regression of octane heats of formation. The associated coefficient of regression and standard error associated with Eq. (2) are $R = 0.934$ and $S = 0.417$. Hence, Eq. (2) will reproduce the heats of formation of octane isomers, on average, with the above standard error. What is unsatisfactory then about such a regression, assuming that we were justified to use seven descriptors $X_1 - X_7$?

The above standard error refers to the "good" part of the regression analysis: its reproduction of the input data. The unsatisfactory part is reflected in the statistics of the individual coefficients of the regression that we list below:

Variable	Coefficient	Standard error of the coefficients
Const.	1.41	94.80
X_1	−10.08	29.01
X_2	−4.86	3.94
X_3	−4.50	7.29
X_4	−0.04	10.52
X_5	−1.34	13.06
X_6	2.01	15.78
X_7	5.83	18.42

Clearly, from the statistical point of view, the regression equation is meaningless, i.e., the coefficients of the equation could not have any possible meaning, since the errors are often several times larger than the coefficients themselves. A referee of a manuscript in which the above regression equation was mentioned lamentably stated:

> In general, the results from this kind of approach are terrible. Correlation coefficients are excellent, computed standard deviations were very small, but when the regression equations were examined, it was found that the standard deviations of the regression coefficients were usually much larger than the values of the coefficients themselves. **It was gradually recognized that such equations were useless for predictive purposes, and statistically invalid for correlating data.**

I would agree with much of the factual statements of the above quote, but strongly disagree with its conclusion, which I have emphasized. This article, in a way, is a response to the incorrect conclusion of the above anonymous critic of multiple

regression analysis. The critic apparently has not recognized the paradoxical aspect of the standard multiple regression analysis. The proper conclusion of the statistical behavior of the coefficients of the above regression equations is that **the equations were useless for (coefficient) interpretation, but their predictive purpose for correlating data is statistically valid.**

Use of the equations and the validity of the predictions made by the equations are not necessarily related to their instability. The validity of the equations is measured by the standard errors s_k of the coefficients; the quality of the predictions made by the equations are determined by the standard deviation S. The latter can be statistically significant and acceptable, whereas the former are at the same time statistically unacceptable. The instability of the equations means that the coefficients of the equation need not be reproducible, but the "predictions" based on such equations, nevertheless, can give statistically acceptable results.

We will support our claim in the next section. Before that, let us briefly discuss the conclusions that the anonymous critic offered. First, if he/she would be right that would have devastating consequences on the work of many people who over the last several decades, if not longer, used multiple regression analysis. Could it be that most of the past structure–property and structure–activity correlations were a waste? That the work of Corwin Hansch and his school [4] was all meaningless? I don't think so.

The critic claims that this apparent fault of the multiple regression "was gradually recognized" as to produce equations that "were useless for . . . correlating data." Recognized by whom? Published where?

What few may have realized over the years is that widely different equations can produce numerically similar results for the prediction of properties. In other words, the regression equations that reproduce molecular properties within a given accuracy (measured by the standard deviation S) need neither to be unique nor even closely similar. Because the coefficients of the equations can be widely different, they cannot possibly have a statistical significance. One may say that the equations of multiple regression are unstable, having elements of a "chaotic" behavior, in that a small perturbation of the equations, such as obtained by introducing an additional variable that only slightly reduces the standard error S, may drastically change the coefficients of the already-present variables.

On Instability of Regression Equations

To illuminate the instability of the regression equations, we listed in Table I the regression equations obtained by a successive stepwise inclusion of additional variables. We will continue to use the connectivity indices χ_1–χ_5 as descriptors and consider application to octane isomers, but the outline and the conclusions are more general. They apply to correlations of other molecules and other properties and, in particular, to correlations using other descriptors. The left half of Table I shows the variables used, the coefficient of the regression equation, and the corresponding standard error for the coefficient. At the right half of the table for each regression equation, we give the coefficient of the regression R, the standard error

TABLE I. The regression equations (column two) and the standard error of the coefficients of the equations (column three) for stepwise inclusion of the descriptors (the connectivity indices χ_k).

Variables	Coefficient	Standard error	R	S	F
Const	**75.2462**	**3.62**	0.679	1.699	41.82
χ_1	**−6.4225**	**0.99**			
Const	16.9912	14.88	0.930	0.489	48.18
χ_1	6.2484	3.27			
χ_2	**3.7949**	**0.95**			
Const	37.410	50.32	0.931	0.503	30.43
χ_1	1.9202	10.70			
χ_2	2.554	3.07			
χ_3	**−0.3829**	**0.90**			
Const	75.6936	69.53	0.935	0.509	22.42
χ_1	−5.918	14.54			
χ_2	0.1327	4.32			
χ_3	−1.2092	1.37			
χ_4	**−0.6376**	**−0.79**			
Const	19.2243	1.15	0.965	0.391	32.46
χ_1	3.8336	11.57			
χ_2	4.2371	3.56			
χ_3	1.5555	1.36			
χ_4	1.8334	0.99			
χ_5	**3.634**	**1.15**			
Const	23.3773	57.61	0.9667	0.398	25.14
χ_1	1.2413	12.30			
χ_2	4.5566	3.65			
χ_3	2.8204	2.20			
χ_4	3.5232	2.49			
χ_5	5.6669	2.98			
χ_6	**3.2272**	**4.20**			
Const	−0.4417	94.37	0.9670	0.416	20.60
χ_1	9.7097	28.88			
χ_2	4.8577	3.92			
χ_3	0.5693	7.25			
χ_4	0.2000	10.48			
χ_5	1.5244	13.00			
χ_6	−1.8233	15.71			
χ_7	**−6.0037**	**18.34**			

R is the coefficient of the regression; S is the standard error of prediction; and F is the Fisher ratio.

S, and the Fisher ratio F. Since all the molecules considered are of a same size, the dominant dependence of the property on the molecular size is constant and the three parameters R, S, and F all represent equally well the quality of the regression. We will follow their variations as the number of descriptors are increased.

The linear correlation, based on a single descriptor,

$$P = -6.4225x_1 + 75.2462, \tag{3}$$

the first equation in Table I, does not yield a satisfactory result: The regression coefficients are too low, reaching barely 0.679. Hence, the equation explains the variance in less than half of the data points (the coefficient of determination $R^2 = 0.461$). When we combine the connectivity indices x_1 and x_2, already we obtain a relatively satisfactory correlation:

$$P = 6.2484x_1 + 3.7949x_2 + 16.9912. \tag{4}$$

The standard error is now reduced by more than three times. The coefficient of the regression is $R = 0.930$ and the Fisher F ratio is significantly increased. If one stops at this stage in the structure–property analysis and tries to interpret the results, i.e., tries to identify structural components that can explain most of variations in the data, one already faces serious difficulty: The two regression equations, one based on x_1 and the other equation using two descriptors x_1 and x_2, bear no resemblance. Consequently, contributions of the individual descriptors cannot be determined unambiguously.

The anonymous criticism: "The results form this kind of approach are terrible," does not apply to the first equation. That equation was unsatisfactory, i.e., not being good enough from a predictive point of view (large S), because the descriptor x_1 did not account for the major part of the isomeric variations. The equation was associated with, relatively speaking, a small standard error. When the second descriptor was introduced, we observed a dramatic drop in the value of the constant term, from 75.24 to 16.99. The coefficient of the initial variable x_1 also shows a "wild" behavior. It shows a similar magnitude but has changed its sign: from -6.42, it became 6.24. The standard error for the constant term has jumped from about 5 to 50% and for the coefficient of x_1 from about 15 to 50%. The standard error for the coefficient of the new term, x_2, is, however, respectable. In contrast, the standard error for "prediction" of the property shows a significant decrease. It dropped from 1.70 to 0.49.

If we include the additional descriptors x_3 and x_4, the new variables did not reduce the standard error, which remained more or less constant. The corresponding coefficients of the regression slightly increased and the F ratio decreased, suggesting limitations of the two descriptors, x_3 and x_4. The additional variables x_3 and x_4 did not introduce significant improvement in the multiple regression analysis, beyond the description already arrived at using x_1 and x_2. However, we are here interested in the mathematics behind the regression equations, so we will continue with the stepwise inclusion of molecular descriptors in order to see how extension of the regression influences the stability of the regression equations.

Consider the values for the constant term in subsequent equations: 75.24; 16.99; 37.41; 75.69; 19.22. So what is the constant term in linear regression of the heats of formation of octanes? The same can be asked about the contribution of x_1, which varies as -6.42; $+6.25$; $+1.92$; -5.92, $+3.83$. The variation of the individual contributions show that something strange is accompanying the multiple regression analysis.

The instability that we observe is a consequence of very strong mutual intercorrelation of the connectivity indices, the descriptors x_1–x_5. Each time that we introduce a new descriptor, since it correlates with the descriptors already used, it modifies their contributions and changes their magnitudes, which have hitherto characterized the roles of the previous variables.

Let us now focus attention on the last coefficient in each of the equations of Table I: the coefficient describing the role of the last variable of the stepwise regression. We claim that this coefficient measures the true role of that descriptor in the regression. We indicated those coefficients in Table I by bold face. They are reproduced here, accompanied with the statistical parameters for the corresponding stepwise equation:

Added variable	Coefficient	R	S	F	New label	
Const	75.2462				Ω_0	
x_1	−6.4225	0.679	1.699	41.8	Ω_1	
x_2	3.7949	0.930	0.489	48.2	Ω_2	(5)
x_3	−0.3829	0.931	0.503	30.4	Ω_3	
x_4	−0.6376	0.935	0.509	22.4	Ω_4	
x_5	3.634	0.965	0.391	32.4	Ω_5	

Furthermore, we claim that the above is the correct regression equation that, first, shows a numerical stability and, consequently, second, allows one to interpret the relative roles of the associated variables. The new variables were designated as Ω_k and represent the corresponding orthogonal components of the initial variable x_k. The claim follows our previous work on orthogonalization of molecular descriptors [5,6]. The new variables Ω_k describe those parts of the stepwise-introduced descriptor that do not correlate with any previously used descriptors. The construction of such descriptors is based on residuals of regressions with the descriptors already used, which, in turn, are based on residuals of their intercorrelations, and so on.

Observe that the R and S values listed above are precisely the same R and S values listed in Table I, yet the approach of Table I has been attacked by the anonymous critic as "terrible" and "useless." However, the predictions based on the equations of Table I are as good as the predictions based on the corresponding truncated equation listed. Hence, since the predictions based on the above equations are valid (*vide infra*), and they do not differ from the predictions based on the equations of Table I, then one concludes that "unstable" equations can equally well give a prediction of properties.

What Table I does not permit is for one to interpret the contributions of individual variables. Hence, in the case of three variables, we can use either

$$P = 37.410 + 1.0202x_1 + 2.554x_2 - 0.3829x_3$$

or

$$P = 75.2462 - 6.4225\Omega_1 + 3.7949\Omega_2 - 0.3829\Omega_3. \tag{6}$$

Both equations will equally well reproduce the input data. Observe, as we have already said, that the last coefficients in such comparative equations will always be same, since they describe the contribution of the new variable.

If we decide that only the first two variables are sufficient to represent the regressions, we have either

$$P = 16.9912 + 6.2484\chi_1 + 3.7949\chi_2$$

or

$$P = 75.2462 - 6.4225\Omega_1 + 3.7949\Omega_2. \tag{7}$$

The first equation had to be constructed, whereas in the case of orthogonalized descriptors, we can truncate the equation in (6) and immediately write down the new equation. Because the truncation process does not change the coefficients if the variables are orthogonal, we can attribute to the first connectivity index Ω_1 or χ_1 the contribution -6.4225, and to the second variable Ω_2, which is that part of χ_2, which does not parallel χ_1, the value $+3.79$, etc. Moreover, if we want to use three descriptors, we could select Ω_1, Ω_2, and Ω_5 as the best among the set of the considered variables and eliminate Ω_3 and Ω_4 from the last equation in Table II to obtain

$$P = 75.2462 - 6.4225\Omega_1 + 3.7949\Omega_2 + 3.634\Omega_5. \tag{8}$$

Statistical Analysis of the Orthogonal Equations

The cause of the instability of the regression equation is interdependence among the variables. Those who did not experience the difficulties of instability of the regression equations were fortunate that their problems did not call for variables that were strongly intercorrelated. However, in structure–property and structure–activity studies, as a rule, one has not only intercorrelated variables but also often very strongly intercorrelated variables—the case usually referred to as collinearity.

We will now substantiate our claim on the numerical stability of the regression equations based on orthogonalized descriptors and justify the *ad hoc* construction of the "stable equations" from the coefficients of stepwise regressions. In Table II, we present the information on the standard errors for the coefficients of the regression equations based on orthogonal descriptors rather than on "unrefined," intercorrelated, connectivity indices $\chi_1-\chi_5$. Table II should be compared to Table I: They both list sets of regression equations by giving the numerical values for the coefficients of all variables and the corresponding standard errors. Since the overall standard error of prediction S and the regression coefficients R are the same, whether descriptors are orthogonal or not, the values of R and S are not repeated in Table II. Neither is the Fisher F ratio showed as it also does not change, despite the numerical stability of the equations.

TABLE II. The regression equations and the standard error of the coefficients of the equations for stepwise inclusion of orthogonalized connectivity indices Ω_k.

Variable	Coefficients	Standard error	F-ratio
Const	75.2462	3.62	41.82
Ω_1	−6.4224	0.99	
Const	75.2454	2.61	48.18
Ω_1	−6.4222	0.72	
Ω_2	3.7949	0.95	
Const	75.2456	2.69	30.43
Ω_1	−6.4222	0.74	
Ω_2	3.7949	0.98	
Ω_3	−0.3833	0.90	
Const	75.2463	2.72	22.43
Ω_1	−6.4224	0.75	
Ω_2	3.7949	0.99	
Ω_3	−0.3833	0.91	
Ω_4	−0.6378	0.79	
Const	75.2467	2.09	32.45
Ω_1	−6.4225	0.57	
Ω_2	−3.7945	0.76	
Ω_3	−0.3836	0.70	
Ω_4	−0.6380	0.60	
Ω_5	3.6317	1.15	

F-ratios (as well as R and S) are the same as in Table I.

First to observe in Table II is that the standard errors of the coefficients are smaller, often much smaller, than the corresponding coefficients. Hence, statistically speaking, the coefficients of the regression equations are significant. The equations of Table II fully negate the concerns of the anonymous critic and those who "gradually recognized that such equations were useless."

There is yet another very interesting and important result that can be seen from Table II. The standard error of a same coefficient in different equations decreases as one introduces additional descriptors. This is just the opposite of what we have seen in Table I, when new nonorthogonal descriptors caused a dramatic increase in the standard errors for the coefficients of the hitherto considered equations. For example, when using orthogonal descriptors, the initial standard error for the constant term 75.2462 is 3.6242 (about 4.8%). After we added new (orthogonal) descriptors, it decreased gradually to 2.61, 2.68, 2.71, and, finally, 2.09 (i.e., 2.8%). If one looks carefully at the entries of Table II, one can see that the decrease in the standard error for the coefficients was not uniform. The increase in the standard error (upon introducing Ω_3 and Ω_4) is also reflected in the small decrease of the F ratio. This suggests a worsening of the quality of the regression from the statistical

TABLE III. The correlation matrix for the connectivity indices (upper part) and orthogonalized
connectivity indices (lower part).

	H_f	χ_1	χ_2	χ_3	χ_4	χ_5
H_f	1.000	−0.500	0.912	−0.207	−0.440	−0.134
χ_1		1.000	−0.976	−0.185	0.527	0.577
χ_2			1.000	−0.023	−0.521	−0.417
χ_3				1.000	−0.452	−0.620
χ_4					1.000	0.125
χ_5						1.000

	H_f	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5
H_f	1.000	−0.850	0.337	−0.041	−0.080	0.240
Ω_1		1.000	0.000	0.000	0.000	0.000
Ω_2			1.000	0.000	0.000	0.000
Ω_3				1.000	0.000	0.000
Ω_4					1.000	0.000
Ω_5						1.000

viewpoint. This variation of the standard error of the coefficients can assist one in selecting optimal descriptors. Based on these considerations, we can suggest as optimal the regression equation

$$P = 75.2467 - 6.4225\Omega_1 + 3.7945\Omega_2 + 3.6317\Omega_5. \qquad (9)$$

Due to the iterating nature of the orthogonalization algorithm, one can have small oscillations of individual coefficients (see Table II). These are due to accumulation of the rounding errors, and with increased precision (double or even higher), these fluctuations are eliminated.[†]

Concluding Remarks

The root of the instability of the multiple regression equation has been a strong interdependence of descriptors. The intercorrelation of the descriptors used is illustrated in the correlation matrix for the connectivity indices (Table III). Observe the strong correlation between χ_1 and χ_2. The index χ_3 is relatively weakly correlated with χ_1 and χ_2, but it appears that χ_3 is not very relevant for the molecular property of octanes, as already discussed. In contrast, the correlation matrix for the orthogonalized descriptors Ω_k shows that the new variables are not intercorrelated (the lower part of Table III). The submatrix describing new intercorrelations is the identity matrix.

[†] D. Plavšić (Institute Rudjer Bošković, Zagreb, Croatia, 1993) carried orthogonalization for the connectivity indices using 20 digit arithmetics and obtained complete agreement for the coefficients on 20 digits.

The purpose of this article was to demonstrate that multiple regression analysis continues to be a viable and important theoretical tool for data reduction. Although most users have not shown openly their concern about the lack of the statistical significance of regression equations, apparently few, who "gradually recognized that such equations were useless for predictive purposes," have been confused. They failed to distinguish between the statistical validity of the predictions and the statistical validity of the equations.

To see the difference, consider an analogy, a comparison between the Copernican planetary system and the geocentric system of Ptolomy: If we are interested in predictions of solar or lunar eclipses, both approaches would give valid predictions. Even the actual amount of computation need not be significantly different. But one should not try (with the exception of the members of Flat Earth Society!) to give physical significance to the equations behind the epicentric orbits of the geo-centric astronomy. The objections to regression analysis amount to the claim that the geocentric system (in astronomy) is useless. But it was useful. It served people well before adoption of the Copernicus heliocentric system. The difference between the two is in substance, or as Kepler replied when confronted with objections to his theory: "It is simpler, even if it cannot explain everything that Ptolomy theory can."

So, the orthogonal descriptors are simpler than are descriptors that are intercor-related, not necessarily computationally, but certainly conceptually. Moreover, even if they cannot always give simple explanations, they can give an explanation nev-ertheless, whereas nonorthogonal descriptors, at best, can offer some ambiguous description for the variables involved.

Acknowledgments

Bibliography

[1] E. Mendelson, *Introduction to Mathematical Logic* (Van Nostrand, Princeton, NJ, 1964), p. 3.
[2] M. Randić, J. Am. Chem. Soc. **97**, 6609 (1975).
[3] L. B. Kier, W. J. Murray, M. Randić, and L. H. Hall, J. Pharm. Chem. **65**, 1806 (1976).
[4] C. Hansch, Acc. Chem. Res. **2**, 232 (1969).
[5] M. Randić, New J. Chem. **15**, 517 (1991).
[6] M. Randić, J. Chem. Inf. Comput. Sci. **31**, 311 (1991).

The Neglect of Diatomic Differential Overlap (NDDO) Fragment Self-Consistent Field Method for the Treatment of Very Large Covalent Systems

GÁBOR NÁRAY-SZABÓ AND GERGELY TÓTH

Department of Theoretical Chemistry, Eötvös University Budapest, H-1518 Budapest 112, P.O. Box 32, Hungary

GYÖRGY G. FERENCZY

Chemical Works of Gedeon Richter Ltd., H-1475 Budapest, P.O. Box 27, Hungary

GÁBOR CSONKA

Technical University of Budapest, H-1111 Budapest, Szt. Gellért tér 4, Hungary

Abstract

We present a semiempirical NDDO procedure, called the fragment SCF (FSCF) method, to treat very large molecules. The covalent system is partitioned into a relatively small subsystem where substantial chemical changes take place and an environment that remains more-or-less unperturbed during the process. We expand the wavefunction on an atomic hybrid basis and perform an SCF procedure for the subsystem in the field of the iteratively determined electronic distribution of the environment. We wrote a program for the IBM RISC/560 computer and did several test calculations for a variety of large classical molecules. Protonation energies, proton transfer potential curves, rotational barriers, atomic net charges, and HOMO and LUMO energies, as computed by the exact version of the NDDO method, are fairly well reproduced by our approximation. Using the FSCF method, we calculated the molecular electrostatic potential on the van der Waals envelopes of the specificity pocket of trypsin and the lysine side chain of the bound substrate and visualised electrostatic complementarity. We developed a novel bulk phase Monte Carlo simulation technique and calculated the energy by the above approximation and applied the method to amorphous silicon (a-Si). Starting from a distorted tetrahedrally bonded random network model of a-Si with 216 atoms, we performed Monte Carlo simulations using the FSCF energy calculation. For the second and subsequent configurations, we exploited the feature of the Metropolis–Teller algorithm, namely, that, to generate a new configuration, we displace only a single atom. Thus the number of integrals to be calculated drastically decreases since only those have to be reevaluated that contain the coordinates of the displaced atom. After equilibration we obtained distribution functions being almost identical to the one corresponding to the distortion free tetrahedrally bonded network. The same technique was applied to liquid chlorosilanes. We found that Si—Cl bonds elongate by 6 to 16 pm while H–Si–Cl and Cl–Si–Cl angles change by 2–4° as compared to the gas phase. © 1994 John Wiley & Sons, Inc.

Introduction

Quantum chemical calculations for small and medium-size organic molecules in the ground state became almost routine at the *ab initio* and semiempirical level of approximation [1,2]. One of the challenges remained is the treatment of very large systems (proteins, nucleic acids, molecular liquids, surfaces, amorphous ma-

terials, etc.) containing hundreds to thousands of atoms. Computer simulations for these systems are based in most cases on empirical force fields [3] that may provide excellent results for a given problem but are, in general, not completely transferable. Accordingly, comparison of results by different softwares is difficult. Calculations for reaction paths are especially problematic and need special treatment in each case. Another problem is that open-shell systems or problems related to charge transfer cannot be handled.

The most popular approach to the molecular orbital theory of very large covalent systems is to combine quantum mechanics, applied to the subsystem where the important chemical changes take place, with molecular mechanics that describes the environment [4–7]. The major problem with this method is the definition of the subsystem–environment boundary where hydrogen atoms have to be put in order to provide a classical closed-shell model of the subsystem for which the quantum mechanical calculation can be done. These hydrogens may provide spurious interaction energy terms with the surrounding atoms and thus lead to erroneous results. In quantum/classical methods, polarization of the environment by the subsystem is treated by introducing empirical parameters. This may yield very good agreement with experiment, e.g., for proteins, but empirical polarizabilities are available only for a limited class of atoms. Therefore, some interesting systems (e.g., silicates, zeolites, nonaqueous solvents) provide special problems. Reaction field theories are also popular; but, in some cases, the solvent model is oversimplified, or we have a nonlinear Schrödinger equation to be solved and may face computational difficulties [8–11].

In order to provide a general and consistent solution to the above problems, we have been working on our fragment self-consistent field (FSCF) method for more than a decade [12–15]. Our philosophy, similar to that outlined in Ref. [4], is to partition the whole covalent system to be treated into a central part (subsystem) where important changes (chemical reaction, conformational change, excitation, ionization, etc.) take place and environment that has only a secondary effect on this localised event. This allows restriction of the sophisticated SCF treatment to the subsystem and introduction of more and more approximations with increasing distance from it. Such a model should account for charge transfer inside the subsystem, polarization between subsystem and close surroundings, and the electrostatic effect of the far-lying environment on the subsystem and close surroundings, respectively.

Derivation of a molecular orbital theory on the basis of the above model is straightforward if we deal with minimum basis sets. In the case of larger than minimal basis sets, there is an ambiguity in the subsystem–environment partition, which may lead to imbalance and spurious results. The success of the semiempirical neglect of diatomic differential overlap (NDDO) methods in the treatment of a wide variety of molecular systems [2] tempted us to apply the FSCF model at the NDDO level both in the AM1 [16,17] and PM3 [14,18] parameterizations. In the following we outline the method then we report on applications to various large covalent systems: proteins, amorphous materials, and liquids.

Method

Basis Set

The same way as we may build up molecules from more or less transferable chemical bonds, we expand the wavefunction on the basis of strictly localized molecular orbitals (SLMOs). For closed-shell classical molecules we define one-center lone pairs, two-center σ-, and many-center π-orbitals [19–23]

$$\phi_i^{\mathrm{LP}} = h_{Ai} , \tag{1}$$

$$\phi_i^{\sigma} = c_{Ai} h_{Ai} + c_{Bi} h_{Bi} , \tag{2}$$

$$\phi_i^{\tau} = \sum_m c_{mi} h_{mi}^{npz} , \tag{3}$$

where

$$h_{Ai} = b_i^{ns} u_A^{ns} + b_i^{npx} u_A^{npx} + b_i^{npy} u_A^{npy} + b_i^{npz} u_A^{npz} \tag{4}$$

is a normalized atomic hybrid orbital (HYO) centered on atom A. HYOs on the same atom are Löwdin-orthogonalized to obtain b_i coefficients in Eq. (4). u_A^{ns}, $\cdots$ are Slater-type orbitals with principal quantum number n. The total wavefunction is the antisymmetrized product of the SLMOs ($2N$ is the number of electrons considered):

$$\Psi = \det|\phi_i^{\alpha}(1)\phi_i^{\beta}(2)\cdots\phi_N^{\alpha}(2N-1)\phi_N^{\beta}(2N)| . \tag{5}$$

For the subsystem we expand the wavefunction on the basis HYOs ($2N_S$ is the number of electrons in the subsystem):

$$\Psi^S = \det|\psi_i^{\alpha}(1)\psi_i^{\beta}(2)\cdots\psi_N^{\alpha}(2N_S-1)\psi_N^{\beta}(2N_S)| \tag{6}$$

with

$$\psi_i = \sum_A \sum_{j\in A} a_{ij} h_{Aj} . \tag{7}$$

Secular Equations

In order to consider polarisation (inductive) effects we may optimize HYO coefficients c_{mi} in Eqs. (2)–(3) to obtain ϕ_i^{opt} (the zeroth-order wave function) by solving a set of coupled 2×2 or $m \times m$ secular equations for SLMOs of the subsystem and close environment

$$F_i c_{mi} = \varepsilon_{mi} c_{mi} . \tag{8}$$

If we assume that the core Hamiltonian $\mathbf{H}$ and the density matrix $\mathbf{P}$ are block diagonal and the differential overlap between HYOs a and b is zero, the Fockian is written as follows (cf. Ref. [14] for more details):

$$F_{aa} = H_{aa}^{\text{MNDO}} + \tfrac{1}{2} P_{aa}(aa|aa) + \sum_{b\,(\neq a)\in\alpha} P_{bb}(aa|bb)$$

$$+ \sum_{\beta(\neq\alpha)} \sum_{c\in\beta} P_{cc}(aa|cc), \qquad a\in\alpha, \quad (9a)$$

$$F_{ab} = H_{ab}^{\text{MNDO}} - \tfrac{1}{2} P_{ab}(aa|bb), \qquad a, b \in \alpha, \quad a \neq b, \qquad (9b)$$

$$F_{ac} = 0, \qquad a \in \alpha, \quad c \in \beta, \quad \alpha \neq \beta. \qquad (9c)$$

Here a, b, c, and d refer to HYOs, $\mathbf{H}^{\text{MNDO}}$ and $\mathbf{P}$ are the core Hamiltonian in the MNDO approximation and the density matrix, $(ab|cd)$ denotes electron interaction integrals in the usual convention, and α and β stand for an SLMO.

For the subsystem we have the following secular equation: (S and E denote the subsystem and the environment)

$$\mathbf{F}^{S}\boldsymbol{a}_i = \varepsilon_i \boldsymbol{a}_i \qquad (10)$$

with

$$F_{ab}^{S} = H_{ab} + \sum_{c,d\in S} P_{cd}[(ab|cd) - \tfrac{1}{2}(ac|bd)], \qquad (11)$$

$$H_{ab} = H_{ab}^{\text{MNDO}} + \sum_{\alpha\in E}\sum_{c\in\alpha} P_{cc}(ab|cc) - \tfrac{1}{2}\sum_{\alpha\in E}\sum_{c,d\in\alpha} P_{cd}(ac|bd). \qquad (12)$$

The dimensionality of Eq. (10) is proportional to the size of S; the number of HYOs in E appears only in the electron interaction terms. Accordingly, the computational work reduces very much.

Monte Carlo Simulations

We applied the Metropolis–Teller algorithm to generate configurations [24] and calculated the potential with the NDDO FSCF method. We applied periodic boundary conditions and performed all calculations corresponding to the "minimal image convention" [25]. A new configuration was accepted or rejected as in the conventional Monte Carlo procedure. For the wave function of the starting configuration, we solved the secular equations in Eqs. (8) and (10). For the second, third, and subsequent configurations we exploited the special feature of the Metropolis–Teller algorithm, namely, that, to generate a new configuration, only a single atom is displaced. Thus, only those integrals have to be reevaluated that contain the coordinates of the displaced atom. Other integrals did not change, and the corresponding energy term was constant. With this simplification the computer time was reduced by a factor of 20.

Although we had no problems for amorphous silicon, it was not so easy to get reliable statistics on liquid chlorosilanes by the classical Monte Carlo method. During our simulations the acceptance ratio decreased in some cases that we had to compensate by reducing the maximal step size for molecular translation. This was not necessary for rotational and intramolecular motions. Accordingly, our statistics for quantities that depend on intermolecular distances were not reliable while it was

quite good for quantities, depending on intramolecular and orientation variables. In order to check the reliability of our simulations, we averaged the distributions of bond lengths, bond angles, and intermolecular orientations over each consecutive 10 configurations of the latest 100 ones. A proof for the appropriate convergence of the above parameters is that we could not detect any difference between the pair, cosine, and orientation distribution functions for these averages, respectively. Simulation temperatures (K), densities (g cm^{-3}), and the number of steps in the Metropolis–Teller algorithm were as follows: SiH$_3$Cl: 160.15, 1.145, 30,000; SiH$_2$Cl$_2$: 160.15, 1.42, 20,000; SiHCl$_3$: 273.15, 1.34, 20,000.

Implementation to the SYBYL Software

Since the input for an FSCF calculation should contain precise and unequivocal information on chemical bonding in the system studied, it is quite complicated and, for large molecules, practically impossible to construct. In order to allow simpler input, we wrote an interface to the SYBYL molecular modeling package [26] that uses the MOL and MOL2 file formats. The file is made of rows each containing atom types, Cartesian coordinates, and bonding information with connectivity numbers. At present we may consider 33 atom types with different coordination numbers, charges and geometries for the following elements: H, Li, C, N, O, F, Na, Al, Si, P, S, Cl, K, Ca, Br, and I. Additionally, it is possible to define hypothetical lone-pair centers and dummy atoms, as well.

We had some problems in the definition of delocalized π-bonds given in Eq. (3). For example, in SYBYL, butadiene appears as two double bonds connected by a single bond while the FSCF method implies a four-center 4π-electron delocalized system. This means that all adjacent carbon atoms with sp^2 hybridization must be included. The implementation of an algorithm for the general case is not obvious because it may lead to an infinite loop with respect to the ring systems. We applied a sieving method to find all connected sp^2 atoms that must be repeated iteratively until no new connected atoms can be found.

The steps of the iterative sieving algorithm are the following:

(a) Sorting bonds of the molecule.
(b) Select two atoms of the first double bond as part of the π-system.
(c) Go through all higher order bonds and check whether one of their atoms is already connected to the ones already selected.
(d) If the answer is yes for an atom, select it.
(e) Repeat the above procedure from (c) until no new atom can be found.
(f) Repeat the above procedure from (b) with the next bond until the last bond is reached.

Triple bonds require a similar treatment after processing all π-bonds.

We implemented the above algorithm successfully; however, some specific considerations are required, and SYBYL atom types must be completed by others. Some examples are deprotonated hydroxy oxygen, protonated aromatic nitrogen, and the oxygens of the carboxylate anion. These latter are treated separately because

of the extra electron added to the π-system containing two atoms of the same type. In cases where the SYBYL molecular modeling software considers an aromatic system as nonaromatic bond types must be corrected manually. An example is pyrrole: If it is represented as a set of two double and three single bonds, the nitrogen atom will have three single bonds and a lone pair and will be considered by the above procedure as nonaromatic. Another problem we faced is that if the type of an atom is changed, its geometry may have to be changed. For example, if we replace an sp^2 nitrogen to an sp^3, one without changing the geometry around it from planar to tetrahedral, it is not possible to generate correct hybrids subsequently.

Applications

Small Model Systems

We investigated the effect of the size of the subsystem on the accuracy of some calculated molecular properties [14]. It was found that the optimal choice of the subsystem is a sphere of 500 pm radius. Deprotonation energies of systems, like $CH_3(CH_2)_{10}COOH$, $CH_3(CH_2)_9CHFCOOH$, the Gly74–Ser75–Ser76–Ser77–Glu-(COOH)78–Lys79–Ile80 fragment of α-chymotrypsin, the $(H_2O)_5$-HOH$(H_2O)_5$, and $(H_2O)_5$HOH hydrogen-bonded chains (dissociating protons are denoted by boldface letters) differ from those calculated for the full system by less than 7.5 and 4.3 kJ/mol on average. The energy curve for the proton transfer from the above heptapeptide fragment to an ammonia molecule near the glutamic acid side chain, as obtained by the FSCF approximation, is almost identical to the exact one. Rotational barriers for the $CH_3(CH_2)_8CH_2$—CH_2COOH and $CH_3(CH_2)_8CH_2$—CH_2COO^- molecules around the indicated C—C bond differ by less than 0.4 kJ/mol from the exact value. A larger deviation, 7.2 kJ/mol, is observed for the energy difference of the above heptapeptide fragment in two different conformations that, however, reduces to 1.6 kJ/mol if the subsystem size increases to 600 pm. Atomic charges for the $CH_3(CH_2)_{10}COOH$ and $(H_2O)_6$ molecules differ by less than 1 millielectrons. Even HOMO and LUMO energies, which are considered as nonlocal properties, are reproduced for the $CH_3(CH_2)_{10}COOH$ molecule within an accuracy of 5 kJ/mol.

Electrostatic Recognition in Trypsin

As an application of the FSCF approximation, we calculated the molecular electrostatic potential on the van der Waals surface of the specificity pocket of trypsin as well as the accommodated Lys side chain of the bound substrate. Coordinates were taken from Protein Data Bank for the trypsin-bovine pancreatic trypsin inhibitor complex [27]. We followed the same procedure as in case of the conventional MNDO, AM1, and PM3 wave functions [28]. We integrated the subroutine for the calculation of molecular electrostatic potentials (MEP) and fields (MEF) into the SYBYL software [26]. The wavefunction, necessary for the calculation of the MEP and MEF, is provided by the FSCF procedure as described above; the coordinates where these quantities have to be calculated can be given either by invoking a text

editor or by definition within SYBYL. The latter method allows one to perform a calculation for a set of points on the molecular van der Waals surface. Fragments of the trypsin–inhibitor complex (subsystem, polarizable, and unpolarizable regions with all surface side chains un-ionized) are displayed in Figure 1. Note that in the present calculation we simply dropped the unpolarizable region since its influence is negligible.

We displayed the MEP on the van der Waals envelope of the Lys side chain of the substrate and the accommodating envelope representing the specificity pocket of the enzyme in color (Fig. 2). The electrostatic complementarity is nicely visualized; negative and positive regions around Lys fit onto those with opposite sign on the specificity pocket envelope.

Amorphous Silicon

We did Monte Carlo simulations with the NDDO FSCF method using the PM3 parameterization [16]. We selected a distorted version of the Wooten model of amorphous silicon with 216 atoms and a number density of 5.005×10^{-7} pm^{-3} at $T = 293.15$ K as the starting configuration [29]. This has only tetracoordinated silicon atoms, i.e., exclusively σ-type SLMOs of Eq. (2) were used to construct the zeroth-order wave function. We distorted this starting configuration with a con-
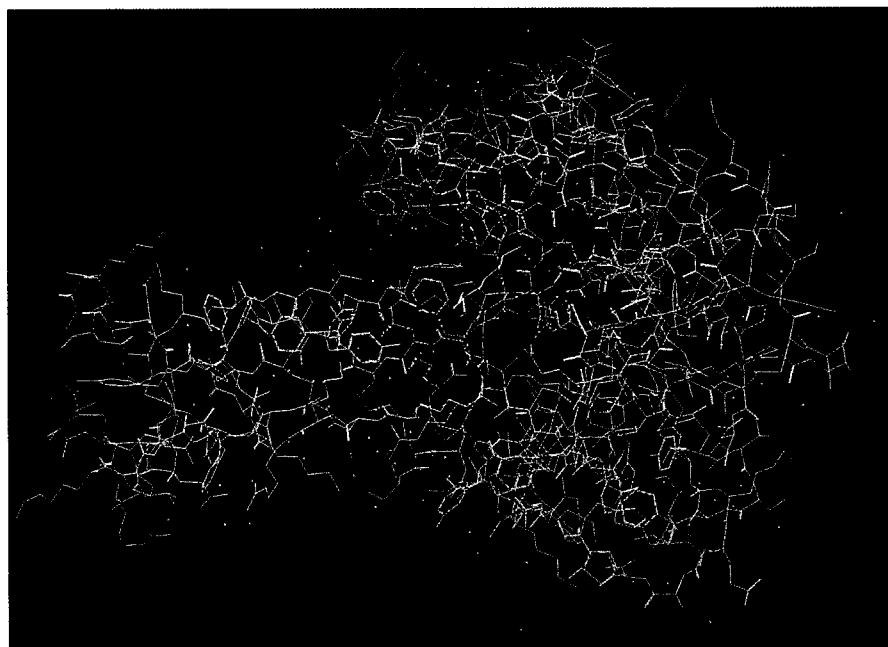

Figure 1.   Computer model of the trypsin-bovine pancreatic trypsin inhibitor complex [27]. Yellow: Lys-15 side chain of inhibitor (subsystem), red: polarizable region, blue: nonpolarizable region.
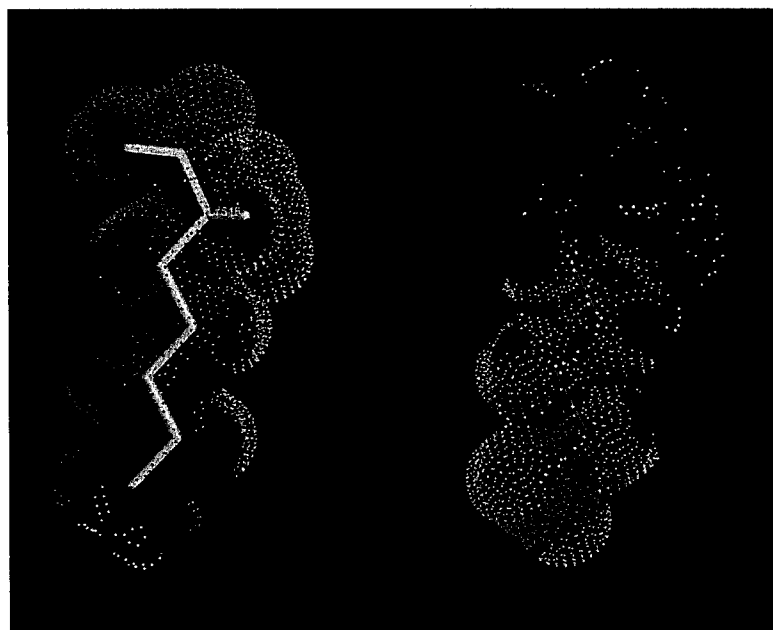
Figure 2.   Molecular electrostatic potential (V) on the van der Waals envelope of Lys-15 emerging from the enzyme (left) and the inhibitor (right). Color codes: $V < 0$ blue, $0 < V < 150$ kJ/mol green, $V > 150$ kJ/mol cyan.

strained hard-sphere Monte Carlo procedure, forcing all Si atoms to have four neighbors within 270 pm distance and Si–Si–Si angles to be larger than 95°.

The subsystem radius was 480 pm including second neighbors of the displaced atom. Reaching an equilibrium after 10,000 successful Monte Carlo steps, we averaged 50 configurations each of which was generated after each 500 steps. The total energy was obtained from the zeroth-order wave function. The radial distribution function fits very well to the original Wooten one (cf. Ref. [29]) with some minor differences in the first neighbor region near 235 pm. This may be due to the failure of the PM3 parameterization to reproduce Si–Si distances precisely [2]. It is important to notice that the radial distribution function is continuous and smooth in the 460–500 pm region, which means that there are no spurious effects at the boundary between the subsystem and the environment. Probably this would not be the case if saturating dangling bonds of the subsystem by hydrogen atoms. The calculated cosine distribution function is also in a good agreement with the original one.

The FSCF MC method allows one to calculate the total energy of the system treated. We obtained 6762 and 6765 kJ/mol/Si atom for our model and the Wooten model, respectively. The slight decrease for the latter is understood on the basis that we did the calculation for room temperature while the Wooten model represents a structure at $T = 0$ K.

TABLE I. Geometry parameters for chlorosilane monomers (upper row) and dimers *in vacuo* (middle row) and in the liquid state (lower row).

Parameter / Molecule	R(Si–Cl) pm		H–Si–Cl < deg		Cl–Si–Cl < deg	
	Ab initio	AM1	*Ab initio*	AM1	*Ab initio*	AM1
SiH₃Cl	217.1	207.1	107.3	109.5	—	—
	218.3	207.3	106.9	109.9	—	—
	—	223	—	107	—	—
SiH₂Cl₂	214.5	206.1	108.1	109.5	109.2	109.6
	215.3	206.2	108.0	109.6	108.4	109.4
	—	214	—	109	—	106
SiHCl₃	212.7	205.0	109.9	109.5	109.0	109.4
	213.2	205.1	110.3	109.7	108.6	109.3
	—	211	—	111	—	107

Liquid Chlorosilanes

We did FSCF Monte Carlo simulations for SiH₃Cl, SiH₂Cl₂, and SiHCl₃ in the liquid phase in order to study the effect of condensation on the molecular geometry [30]. We wanted to separate geometry distortions due to simple hydrogen bonding from bulk effects; therefore, we did calculations for various hydrogen-bonded dimers both with *ab initio* and semiempirical methods. *Ab initio* calculations were carried out at the Hartree–Fock SCF level with the GAUSSIAN 92 program package [31]. The Dunning–Huzinaga valence double-zeta basis set [32] was chosen for the valence electrons, and in the case of Si and Cl the effective core potential of Hay and Wadt were applied [33]. The distance of the two atoms forming the hydrogen bond in the dimer were fixed at 200 pm, and these atoms together with the two silicon atoms were in a collinear position. For the same geometry the calculations were repeated also by the AM3 parameterization [17]. Results are shown in Table I. The trends are in all cases similar both for the *ab initio* and semiempirical calculations, but the magnitude of the changes are smaller in the latter case. Silicon–chlorine distances increase upon hydrogen bonding in all three molecules but much less if applying the AM1 method. However, the trend becomes quite pronounced in the liquid phase for which bulk effects are responsible. The distortion of bond angles is not that large; here again bulk effects have more important impact. Further details are given in Ref. [30].

Acknowledgments

Bibliography

[1] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).
[2] J. J. P. Stewart, J. Comput. Aided Mol. Des. **4**, 1 (1990).
[3] E. R. Davidson, Ed., *Molecular Mechanics and Modelling,* Chem. Rev. **93**, 2337–2582 (1993).
[4] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[5] A. Warshel and S. Russell, Q. Rev. Biophys. **17**, 283 (1984).
[6] P. A. Bash, M. J. Field, and M. Karplus, J. Am. Chem. Soc. **109**, 8192 (1987).
[7] J. Gao, J. Phys. Chem. **96**, 537 (1992).
[8] O. Tapia, F. Sussman, and E. Poulain, J. Theor. Biol. **71**, 49 (1978).
[9] S. Miertus, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).
[10] B. T. Thole and P. T. Van Duijnen, Chem. Phys. **71**, 211 (1982).
[11] D. Rinaldi, M. F. Ruiz-Lopez, and J. L. Rivail, J. Chem. Phys. **78**, 834 (1983).
[12] G. Náray-Szabó and P. R. Surján, Chem. Phys. Lett. **96**, 499 (1983).
[13] G. Náray-Szabó, Croat. Chem. Acta **57**, 901 (1984).
[14] G. G. Ferenczy, J. L. Rivail, P. R. Surján, and G. Náray-Szabó, J. Comput. Chem. **13**, 830 (1992).
[15] G. Náray-Szabó and G. G. Ferenczy, J. Mol. Struct. THEOCHEM **261**, 55 (1992).
[16] G. Tóth and G. Náray-Szabó, J. Chem. Phys. **100**, 3742 (1994).
[17] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 (1985).
[18] J. J. P. Stewart, J. Comput. Chem. **10**, 209 (1989).
[19] G. G. Hall, Proc. Roy. Soc. London, Ser. A**205**, 541 (1951).
[20] C. Sandorfy, Canad. J. Chem. **33**, 1337 (1955).
[21] R. Hoyland, J. Am. Chem. Soc. **90**, 2227.
[22] G. Náray-Szabó, Acta Chim. Acad. Sci. Hung. **40**, 261 (1976).
[23] P. R. Surján, M. Révész, and I. Mayer, JCS Faraday Trans. 2 **77**, 1129 (1981).
[24] N. Metropolis, A. W. Metropolis, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[25] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[26] SYBYL Molecular Modelling Software, Version 6.0a, (TRIPOS Associates, St. Louis, MO, February 1993).
[27] R. Huber and J. Deisenhofer, Protein Data Bank Entry 2PTC, Version of April 1992.
[28] G. G. Ferenczy, C. A. Reynolds, and W. G. Richards, J. Comput. Chem. **11**, 159 (1990).
[29] F. Wooten and D. Wearie, Solid State Phys. **40**, 1 (1987).
[30] G. Tóth, O. Gereben, and G. Náray-Szabó, J. Mol. Struct. THEOCHEM, to appear.
[31] M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzales, R. L. Martin, D. J. Fox, D. J. DeFrees, J. Baker, J. J. P. Stewart, and J. A. Pople, GAUSSIAN 92 (Gaussian Inc., Pittsburgh, PA, 1992).
[32] T. H. Dunning and P. J. Hay, *Modern Theoretical Chemistry* (Plenum, New York, 1976); S. Huzinaga, J. Chem. Phys. **42**, 1293 (1965).
[33] P. J. Hay and W. R. Wadt, J. Chem. Phys. **85**, 270, 284, 299 (1985).

1994 Sanibel Symposia List of Participants

IOURI ABACHKINE
Program Resources, Inc.
National Cancer Institute
Frederick Cancer Research & Development
   Center, P.O. Box B
Frederick, MD 21702

GUSTAVO ARTECA
Laurentian University
Chemistry & Biochemistry Depts
Ramsey Lake Road
Sudbury, Ontario P3E 2C6
CANADA

OSMAN ATABEK
CNRS, Universite de Paris Laboratoire de
   Photophysique Moleculaire
Bat. 213, Campus D'Orsay
Orsay 91405
FRANCE

RICHARD BADER
McMaster University
Chemistry Department
1280 Main St. W.
Hamilton, Ontario L8S 4M1
CANADA

KYOUNG BAECK
University of Florida
Quantum Theory Project
377 WM Hall
Gainesville, FL 32611

JON BAKER
BIOSYM Technologies, Inc.
9685 Scranton Road
San Diego, CA 92121-2777

ANNA BALKOVA
University of Florida
Quantum Theory Project
370 WM Hall
Gainesville, FL 32611

RODNEY BARTLETT
University of Florida
Quantum Theory Project
381 WM Hall
Gainesville, FL 32611

INDER BATRA
IBM Almaden Research Center
K44/802 650 Harry Road
San Jose, CA 95120-6099

SULLIVAN BECK
University of Florida
Quantum Theory Project
345 WM Hall
Gainesville, FL 32611

AXEL BECKE
Queen's University
Department of Chemistry
Kingston, Ontario K7L 3N6
CANADA

DARIO BEKSIC
University of Florida
Quantum Theory Project
348 WM Hall
Gainesville, FL 32611

MICHAEL BENSON
Memphis State University
Department of Chemistry
Memphis, TN 38152

MARK BERGER
Silicon Graphics
1100 Abernathy Rd NE STE.1120
Atlanta, GA 30328

GORAN BERGSON
Uppsala University
Department of Chemistry
P.O. Box 531
Uppsala S-751 21
SWEDEN

MICHAEL BERMAN
Scientific Research Air Force Office of AFOSR/
    NC Suite B115
110 Duncan Ave.
Bolling AFB, DC 20332-0001

MARGARITA BERNAL-URUCHURTU
Estado de Morelos Universidad Autonoma del
Av. Universidad 1001
Col. Chamilpa
Cuernavaca, Morelos
MEXICO

DAVID BERNHOLDT
Pacific Northwest Laboratory
Molecular Science Research Ctr
P.O. Box 999
Richland, WA 99352-0999

MANUEL BERRONDO
Brigham Young University
Physics Department
296-ESC
Provo, UT 84602-1022

DAVID BEVERIDGE
Wesleyan University
Chemistry Department
Hall-Atwater Laboratories
Middletown, CT 06457

RICARDO BICCA DE ALENCASTRO
Rio de Janeiro Universidade Federal de Instituto
    de Quimica
Cidade Universitaria, CT, BL A
Rio de Janeiro, RJ 21941
BRAZIL

DAVID BISHOP
University of Ottawa
Chemistry Department
Ottawa KIN6N5
CANADA

RAYMOND BISHOP
University of Manchester Institute of Science &
    Tech.
Department of Mathematics
P.O. Box 88
Manchester M60 1QD
UNITED KINGDOM

RUUD BLOCK
University of Amsterdam
Physics Chemistry Department
Nieuwe Achtergracht 126
Amsterdam 1018 WS
THE NETHERLANDS

JONATHAN BOETTGER
Los Alamos National Laboratory
Group T-1, MS B221
Los Alamos, NM 87545-0000

EDWARD BOUDREAUX
University of New Orleans
Chemistry Department
Lakefront
New Orleans, LA 70148-2820

J. L. BREDAS
University of Mons
Chimie des Materiaux Nouveaux
Place du Parc 20
Mons B-7000
BELGIUM

MARCUS BREWSTER
University of Florida
Pharmos Corporation
Two Innovation Drive, Suite A
Alachua, FL 32615

ANDERS BROO
University of Florida
Quantum Theory Project
380 Williamson Hall
Gainesville, FL 32611

PAUL BRUMER
University of Toronto
Department of Chemistry
80 St. George Street
Toronto, Ontario M5S 1A1
CANADA

JEAN-LOUIS CALAIS
University of Uppsala
Quantum Chemistry Group
Box 518
Uppsala S-75120
SWEDEN

SYLVIO CANUTO
Universidade de São Paulo
Instituto de Fisica
Cxp 20516
São Paulo, São Paulo 01452-990
BRAZIL

CRISTIAN CARDENAS
University of Florida
Quantum Theory Project
345 Williamson Hall
P.O. Box 118435
Gainesville, FL 32611-8435

MIGUEL CASTRO
University of Montreal
Department of Chemistry
CP 6128, Succursale A
Montreal, Quebec H3C 3J7
CANADA

MAURO CAUSA
University of Torino
Dept. of Inorganic Chemistry
Via P. Giuria 5
Torino 10125
ITALY

ANDRES CEDILLO
Universidad Autonoma Metropolitana
Depto. de Quimica
Apartado Postal 86-219
Villa Coapa, DF 14391
MEXICO

CARY CHABALOWSKI
US Army Research Lab
AMSRL-WT-PC
Aberdeen Proving Ground, MD 21005-5066

PIERRE CHABRIER
University of Florida
Quantum Theory Project
Gainesville, FL 32611

ANNE CHAKA
The Lubrizol Corporation
29400 Lakeland Blvd.
Research Division
Wickliffe, OH 44092-2298

GRZEGORZ CHALASINSKI
University of Warsaw
Chemistry Department
ul. Pasteura 1
Warsaw 02-093
POLAND

MATT CHALLACOMBE
Florida State University
Chemistry Department
Tallahassee, FL 32306-3006

DAVID CHANDLER
University of California
Department of Chemistry
Berkeley, CA 94720

DANIEL CHIPMAN
University of Notre Dame
Radiation Laboratory
Notre Dame, IN 46556-0579

SAN-YAN CHU
National Tsing Hua University
Department of Chemistry
Hsinchu, Tawain 30043
CHINA

PAUL CHUN
University of Florida J. Hillis Miller Health Ctr.
Biochem. & Mol. Biology Dept.
Box 100245
Gainesville, FL 32610-0245

JERZY CIOSLOWSKI
The Florida State University
Department of Chemistry, B-164
Tallahassee, FL 32306-3006

OAKLEY CRAWFORD
Oak Ridge National Laboratory
Health Sciences Research Div.
P.O. Box 2008, NS 6123
Oak Ridge, TN 37831-6123

PETER CSAVINSZKY
University of Maine
Physics and Astronomy Dept.
Orono, ME 04469-0108

CHRISTINE CULOT
University of Namur
Department of Chemistry
61 Rue de Bruxelles
Namur B-5000
BELGIUM

TOM CUNDARI
Memphis State University
Department of Chemistry
3744 Walker
Memphis, TN 38152

JOAQUIM DA MOTTA NETO
University of Florida
Quantum Theory Project
348 WM Hall
Gainesville, FL 32611

HERBERT DACOSTA
University of Florida
Quantum Theory Project
372 WM Hall
Gainesville, FL 32611

ROBERT DAMBURG
Latvian Academy of Sciences
Institute of Physics
Salaspils, Riga 229021
LATVIA

ROBIN DAVIES
University of Wales College of Cardiff
Welsh School of Pharmacy
King Edward VII Avenue
Cardiff, Wales CF1 3XF
UNITED KINGDOM

MURRAY DAW
Sandia National Laboratory
MS9161, Sandia Labs
Livermore, CA 94551

ALEX DE VRIES
Rijksuniversiteit Groningen
Department of Chemistry
Nijenborgh 4
Groningen, AG 9747
THE NETHERLANDS

ALAIN DEDIEU
Universite Louis Pasteur CNRS
Lab. de Chimie Quantique
4 rue Blaise Pascal
Strasbourg F-67000
FRANCE

GERARDO DELGADO-BARRIO
CSIC Instituto de Matematicas y
   Fisica Fundamental
Calle Serrano, 113-123
Madrid 28006
SPAIN

JOSEPH DELHALLE
Universitaires Notre-Dame de la Paix
Departement de Chimie
Rue de Bruxelles 61
Namur B-5000
BELGIUM

BRUCE DERBY
Sun Microsystems Marketing Soni Strang
500 Cypress Creek Rd. W. #250
Ft. Lauderdale, FL 33309

ERIK DEUMENS
University of Florida
Quantum Theory Project
359 WM Hall
Gainesville, FL 32611

KEN DILL
University of California
Pharmaceutical Chemistry, S926
San Francisco, CA 94143-1204

AGUSTIN DIZ
University of Florida
Quantum Theory Project
372 WM Hall
Gainesville, FL 32611

ANTHONY DRIBBEN
University of Florida
Department of Chemistry
305 CRB
Gainesville, FL 32611

CLIFFORD DYKSTRA
Indiana University
Department of Chemistry
402 N. Blackford Street
Indianapolis, IN 46202-3274

RON ELBER
University of Illinois
Chemistry M/C 111
829 W. Taylor
Chicago, IL 60680

EPHRAIM ELIAV
Tel Aviv University
School of Chemistry
Tel Aviv 69978
ISRAEL

THOMAS ENEVOLDSEN
Odense University
Dept. of Chemistry
Campusvej 55
Odense M DK-5230
DENMARK

GUILLERMINA ESTIU
Universidad Nacional de La Plata
Programma Quinor
47 y 115, CC 962
La Plata 1900
ARGENTINA

JACK FAJER
Brookhaven National Laboratory
Department of Applied Science
Upton, NY 11973-5000

ADELBERTO FAZZIO
Universidade de São Paulo
Instituto de Fisica
CxP 20516
São Paulo, SP 01452-990
BRASIL

FRANCISCO FERNANDEZ
La Plata Universidad Nacional de
Departamento de Quimica
Calle 47 esquina 115, C.C. 962
La Plata 1900
ARGENTINA

BERND FEYEN
University of Antwerp
Dep. WIS-INF
Groenenborgerlaan 171
Antwerpen 2020
BELGIUM

SIGHART FISCHER
Technische Universitat Munchen
Physik Department T38
Boltzmannstrasse
Garching 8046
GERMANY

JOSEPH FRIPIAT
Universitaires Notre-Dame de la Paix
Departement de Chimie
Rue de Bruxelles, 61
Namur B-5000
BELGIUM

ED GAINES
Silicon Graphics
1100 Abernathy Rd.
NE.-Ste. 1120
Atlanta, GA 30328

VLADIMIR GALKIN
Ukrainian Academy of Science
Inst. for Low Temperature Physics &
    Engineering,
47 Lenin pr.
Kharkov 310164
UKRAINE

JURGEN GAUSS
Universitat Karlsruhe
Lehrstuhl fur Theoretische y
Physikalische Chemie
Karlsruhe D-76128
GERMANY

DENNIS GERSON
IBM
1503 LBJ Freeway
Dallas, TX 75234

LIONEL GOODMAN
Rutgers University
Chemistry Department
New Brunswick, NJ 08903

MARK GORDON
Iowa State University
Department of Chemistry
Wilhelm 304
Ames, IA 50011-3111

JURGEN GRAFENSTEIN
Max Planck Institute
Physik Komplexes Systeme
Postfach 80 06 65
Stuttgart D-70506
GERMANY

FRITZ GREIN
University of New Brunswick
Chemistry Department
Fredericton, NB E3B 6E2
CANADA

J. GRIGERA
Universidad Nacional de La Plata
CONICET, IFLYSIB
C.C. 67
La Plata 1900
ARGENTINA

STEVEN GWALTNEY
University of Florida
Quantum Theory Project
345 Williamson Hall
Gainesville, FL 32611

MARK HAGMANN
Florida International University
Department of Electrical and Computer
    Engineering
Miami, FL 33199

MICHAEL HALL
Texas A&M University
Department of Chemistry
College Station, TX 77843-3255

NICHOLAS HANDY
University of Cambridge
Department of Chemistry
Lensfield Road
Cambridge CB2 1EW
UNITED KINGDOM

AAGE HANSEN
University of Copenhagen
Department of Chemistry
Universitetsparken 5
Copenhagen 0 DK-2100
DENMARK

DAN HARRIS
Molecular Research Institute
845 Page Mill Road
Palo Alto, CA 94304

FRANK HARRIS
University of Utah
Department of Chemistry
Salt Lake City, UT 84112

MARTIN HEAD-GORDON
University of California
Department of Chemistry
Berkeley, CA 94720-9989

KAI-MING HO
Iowa State University
Department of Physics
Ames Laboratory
Ames, Iowa 50011-3020

GERMUND HOJER
de Mexico Universidad Nacional Autonoma
Facultad de Quimica
Ciudad Universitaria
Mexico, D.F. 04510
MEXICO

YA-WEN HSIAO
University of Florida
Department of Chemistry
226 Lei Hall
Gainesville, FL 32611-2046

CHING HAN HU
The University of Georgia CCQC
1001 Cedar Street
Athens, GA 30602-2556

MING-JU HUANG
University of Florida Center for Drug Discovery
College of Pharmacy
Box 100497
Gainesville, FL 32610

KONCAY HUSEYIN
University of Waterloo
Systems Designs Eng.
Waterloo, Ontario N2L 3G1
CANADA

TSUTOMU IKEGAMI
Keio University
Science & Technology Dept.
3-14-1 Hiyoshi
Kohuku-ku, Yokohama 223
JAPAN

I. A. IL'ICHEVA
Russian Academy of Sciences
Institute of Molecular Biology
Vavilov Str. 32
Moscow B-334
RUSSIA 117984

YASUYUKI ISHIKAWA
University of Puerto Rico
Department of Chemistry
P.O. Box 23346
San Juan, PR 00931-3346

SAUL JACCHIERI
Minas Gerais Universidade Federal de
   Depto. de Bioquimica
Pampulha
Belo Horizonte, MG 31270
BRAZIL

MEHRAN JALAIE
Purdue University
Department of Chemistry
402 North Blackford Street
Indianapolis, IN 46202

JAN JENSEN
Iowa State University
Chemistry Department
Gilman Hall
Ames, IA 50011-3111

BOGUMIL JEZIORSKI
University of Warsaw
Chemistry Department
Pasteura 1
Warsaw 02-093
POLAND

BENNY JOHNSON
Q-Chem, Inc.
7520 Graymore Road
Pittsburgh, PA 15221

HERBERT JONES
Florida A&M University
Physics Department
Tallahassee, FL 32307

POUL JORGENSEN
Aarhus University
Department of Chemistry
Langelandsgade 140
Aarhus C DK-8000
DENMARK

UZI KALDOR
Tel Aviv University
School of Chemistry
Tel Aviv 69978
ISRAEL

MICHAEL KASHA
Florida State University
Institute of Molecular Biophysics
Tallahassee, FL 32306-3015

PATRICK KELLY
John Wiley & Sons, Inc.
605 Third Ave.
New York, NY 10158-0012

JAMES KING
U.S. Army Edgewood RD&E Center
SCBRD-RTC
Aberdeen Proving Ground, MD 21010-5423

KARL KIRSCHNER
Lake Forest College
Chemistry Department
555 N. Sheridan
Lake Forest, IL 60045

BERNARD KIRTMAN
University of California
Department of Chemistry
Santa Barbara, CA 93106-9510

DOUGLAS KLEIN
Texas A&M University
Marine Science Department
Galveston, TX 77553-1675

ANTONIOS KOURES
Allied Signal
50 E. Algonquin Rd.
Des Plaines, IL 60017-5016

MORRIS KRAUSS
NIST Center for Advanced Research
  Biotechnology
9600 Gudelsky Drive
Rockville, MD 20850

CARLOS KUBLI-GARFIAS
University of Mexico
Biomedical Institute
Apartado Postal 70-469
Mexico, DF
MEXICO

HENRY KURTZ
Memphis State University
Department of Chemistry
Memphis, TN 38152

JANOS LADIK
Erlangen-Nuremberg University
Institut fur Theoretische Chemistry
Egerlandstrasse 3
Erlangen D-91058
GERMANY

ANTONIO LAGANA
Universita di Perugia
Dipartimento di Chimica
Perugia 06100
ITALY

WILLIAM LAIDIG
Proctor & Gamble Company
Miami Valley Laboratories
P.O. Box 398707
Cincinnati, OH 45239-8707

UZI LANDMAN
Georgia Institute of Technology
School of Physics
Atlanta, GA 30332

GUY LARSON
Brigham Young University Physics &
  Astronomy
296 ESC
Provo, UT 84602

WALTER LAUDERDALE
Laboratory Frank J. Seiler Research
2354 Vandenberg Drive
Suite 2A35
USAF Academy, CO 80840-6272

PIERRE LEBRETON
The University of Illinois
Chemistry Department
845 W. Taylor St., M/C 111
Chicago, IL 60607

KEE LEE
WonKwang University
Department of Chemistry
344-2 ShinYong-Dong
Iri 570-749
KOREA

THERESA LEE
U.S. Army Corps of Engineers WES
IM-MI-C, Bldg. 112, Rm. 113
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

YOON LEE
KAIST
Department of Chemistry
Taejon 305-701
KOREA

WILLIAM LESTER
University of California
Department of Chemistry
Berkeley, CA 94720

JERZY LESZCZYNSKI
Jackson State University
Chemistry Department
1400 Lynch Street
Jackson, MS 39217

ZACHARY LEVINE
Ohio State University
Department of Physics
174 W. 18th Avenue
Columbus, OH 43210-1106

ZE-SHENG LI
Jilin University Inst. of Theoretical Chemistry
Changchun 130021
CHINA

QUN LIN
Florida State University
Chemistry Department
Tallahassee, FL 32306-3006

JAN LINDERBERG
Aarhus University
Chemistry Department
Langelandsgade 140
Aarhus C DK-8000
DENMARK

ANTONIO LINO
de Uberlandia Universidade Federal
Depto. de Ciencias Fisicas
Campus Santa Monica
Uberlandia, MG 38400-902
BRAZIL

RICARDO LONGO
University of Florida
Quantum Theory Project
372 Williamson Hall
Gainesville, FL 32611

PER-OLOV LÖWDIN
University of Florida Quantum Theory Project
P.O. Box 118435
362 Williamson Hall
Gainesville, FL 32611-8435

CHI-HAO LUAN
at Birmingham University of Alabama
Laboratory of Molecular
Biophysics, VH 300
Birmingham, AL 35294-0019

STACY LUCAS
University of North Florida
Natural Sciences
4567 St. Johns Bluff Rd. S.
Jacksonville, FL 32224-2645

DAVID MAGERS
Mississippi College
Chemistry Department
Box 4065
Clinton, MS 39058

PER-AKE MALMQUIST
University of Lund Chemical Center
Theoretical Chemistry Dept.
P.O. Box 124
Lund S-22100
SWEDEN

NORMAN MARCH
Oxford University
Theoretical Chemistry Dept.
5 South Parks Road
Oxford, England OX1 3UB
UNITED KINGDOM

MARTIN MARTINOV
Florida State University
Chemistry Department
Tallahassee, FL 32306-3006

RENEE MATTIE
15 Willow Ct.
Downington, PA 19335

CHANGJIANG MEI
University of Florida
Quantum Theory Project
362 Williamson Hall
Gainesville, FL 32611

HORIA METIU
University of California
Department of Chemistry
Santa Barbara, CA 93106-9510

SARA MEZA-HOJER
de Mexico Universidad Nacional Autonoma
Facultad de Quimica
Cuidad Universitaria
Mexico, D.F. 04510
MEXICO

DAVID MICHA
University of Florida
Quantum Theory Project
356 Williamson Hall
Gainesville, FL 32611

FREDERICK MIES
N.I.S.T.
Physics Bldg., Rm. B268
Gaithersburg, MD 20899

MANOJ MISHRA
Indian Institute of Technology
Department of Chemistry
Powai, Bombay 400076
INDIA

STACEY MIXON
Florida State University
Chemistry Department
Tallahassee, FL 32306-3006

BENNY MOGENSEN
University of Florida
Quantum Theory Project
372 Williamson Hall
Gainesville, FL 32611

HENDRIK MONKHORST
University of Florida
Quantum Theory Project
361 Williamson Hall, Box 8435
Gainesville, FL 32611-8435

MARK MONTOYA
University of Florida
Depart. of Chem. Engineering
Gainesville, FL 32611

JORGE MORALES
University of Florida
Quantum Theory Project
362 Williamson Hall, P.O. Box 118435
Gainesville, FL 32611-8435

JOHN MORGAN III
Institute for Advanced Study
School of Historical Studies
Princeton, NJ 08540

ROBERT MORRISON
East Carolina University
Chemistry Department
Greenville, NC 27858-4353

BENGT NAGEL
Royal Institute of Technology
Theoretical Physics Department
Stockholm S-100 44
SWEDEN

GABOR NARAY-SZABO
Eotvos University
Theoretical Chemistry Dept.
P.O. Box 32
Budapest H-1518
BUDAPEST

ROBERT NESBET
IBM Almaden Research Center
K43/802, 650 Harry Road
San Jose, CA 95120-6099

MARSHALL NEWTON
Brookhaven National Lab.
Department of Chemistry
555A, P.O. Box 5000
Upton, NY 11973

JAN NOBEL
University of Florida
Quantum Theory Project
375 Williamson Hall
Gainesville, FL 32611

MARCEL NOOIJEN
University of Florida
Quantum Theory Project
371 Williamson Hall
Gainesville, FL 32611

JAMES NORRIS
Argonne National Lab.
Chemistry Department
9700 Cass Avenue
Argonne, IL 60439

MARCO NUNEZ
Metropolitana Iztapalapa Universidad
   Autonoma
Depto. de Fisica
Apartado Postal 55-534
Mexico, DF 09340
MEXICO

N. YNGVE ÖHRN
University of Florida
Quantum Theory Project
363 Williamson Hall
Gainesville, FL 32611

NEVIN OLIPHANT
University of Florida
Quantum Theory Project
377 Williamson Hall
Gainesville, FL 32611

JUAN ORIERO
University of Florida
Quantum Theory Project
348 Williamson Hall
Gainesville, FL 32611

VINCENT ORTIZ
University of New Mexico
Chemistry Department
Albuquerque, NM 87131

RADOVAN PADJEN
CNET-France Telecom
Laboratoire de Bagneux
196 Ave. Henri Ravers, BP 107
Bagneux, Cedex 92225
FRANCE

ALEJANDRO PALMA
Universidad Autonoma de Puebla
Instituto de Fisicas
y Astrofisica
Tonantzintla, Puebla 72000
MEXICO

CAROL PARISH
Indiana University
Department of Chemistry
402 N. Blackford Street
Indianapolis, IN 46202

WILLIAM PARKINSON
Southeastern LA University
Department of Chemistry
Hammond, LA 70402

AJITH PERERA
University of Florida
Quantum Theory Project
345 Williamson Hall
Gainesville, FL 32611

JORGE PEREZ
de Rio Cuarto Universidad Nacional
Chemistry & Physics Department
Estafeta Postal No. 9
Rio Cuarto 5800
ARGENTINA

GEORGE PERKINS
Sun Microsystems Marketing Soni Strang
500 Cypress Creek Rd. W. #250
Ft. Lauderdale, FL 33309

JOAKIM PERSSON
University of Lund Chemical Center
Theoretical Chemistry Dept.
P.O. Box 124
Lund S-22100
SWEDEN

EMIL POP
Pharmos Corporation
2 Innovation Drive
Alachua, FL 32615

JOHN POPLE
Northwestern University
Chemistry Department
2145 Sheridan Road
Wilmette, IL 60208

PARAS PRASAD
State University of New York
Department of Chemistry
113 Acheson Hall
Buffalo, NY 14214

MICHAEL PROBST
Innsbruck University
Institut fur Allgerneine,
Anorganische und Theor. Chemie
Innrain, Innsbruck
AUSTRIA

GEORGE PURVIS III
CAChe Scientific, Inc.
P.O. Box 5000, MS 13-400
Beaverton, OR 97077

JAMES RABINOWITZ
Agency U.S. Environmental Protection
GTD/HERL, MD-68
Research Triangle Park, NC 27711

HERSCHEL RABITZ
Princeton University
Department of Chemistry
Frick Laboratory
Princeton, NJ 08544-1009

MICHAEL RAMEK
Technische Universitat Graz
Physikal Chem.
Graz, Austria A-8010
EUROPE

MILAN RANDIC
Drake University
Mathematics & Computer Science
Des Moines, IA 50311

ARVI RAUK
The University of Calgary
Department of Chemistry
2500 University Drive, N.W.
Calgary, Alberta T2N 1N4
CANADA

JOSE RECAMIER
de Mexico Universidad de Autonoma
Instituto de Fisica
Lab. de Cuernavaca, AP 139-13
Cuernavaca, Morelos 62191
MEXICO

ALISTAIR RENDELL
SERC Daresbury Laboratory
Theory & Computational
Science Division
Daresbury, Warrington WA44AD
UNITED KINGDOM

PETER REYNOLDS
Office of Naval Research
Physics Division, Code 3121
800 N. Quincy Street
Arlington, VA 22217

WILLIAM RHODES
Florida State University
Chemistry Department
Tallassee, FL 32306-3006

ANN RICHARD
Agency U.S. Environmental Protection
Maildrop-68
Research Triangle Park, NC 27711

NIGEL RICHARDS
University of Florida
Department of Chemistry
301 CRB
Gainesville, FL 32611

BJORN ROOS
University of Lund
Theoretical Chemistry Dept.
Chemical Centre, POB 124
Lund S-221 00
SWEDEN

NOTKER ROSCH
Technische Universitat Munchen
Chemistry, Dept of Theor.
Lichtenbergstrasse 4
Garching D8046
GERMANY

JOHN ROSENBERG
University of Pittsburgh
Dept. of Biological Sciences
Pittsburgh, PA 15260

PIOTR ROZYCZKO
University of Florida
Quantum Theory Project
345 Williamson Hall
Gainesville, FL 32611

JURI RUBIN
Ukrianian Academy of Sciences
Inst. of Low Temperature Phys. & Engineering,
    47, Lenin Ave.
Khazkov 310164
UKRAINE

KEITH RUNGE
University of Florida
Quantum Theory Project
376 Williamson Hall
Gainesville, FL 32611

JOHN SABIN
University of Florida
Quantum Theory Project
355 Williamson Hall
Gainesville, FL 32611-8435

LYNN SALEMI
University of Florida
Quantum Theory Project
345 Williamson Hall
Gainesville, FL 32611

MALUISA SAN ROMAN Z
Edo. Morelos Universidad Autonoma del
Lab. Ingenieria Molecular
Av. Universidad S/N
Cuernavaca, Morelos
MEXICO

RUBEN SANTIAGO ACOSTA
de Mexico Universidad Nacional Autonoma
Instituto de Ciencias Nucleare
Circuito Exterior, AP 70-543
Mexico, DF 04510
MEXICO

HAROLD SCHERAGA
Cornell University
Baker Lab of Chemistry
Ithaca, NY 14853-1301

MICHAEL SCHMIDT
Iowa State University
Chemistry Department
311 Wilhelm Hall
Ames, IA 50011-3111

PETER SCHMIDT
Office of Naval Research Chemistry Division
800 N. Quincy Street
Arlington, VA 22217-5000

DIETER SCHUCH
J.W. Goethe-Universitat
Inst. fur Theoretische Physik
Robert-Mayer Str. 8-10
Frankfurt am Main 60054
GERMANY

GUSTAVO SCUSERIA
Rice University
Chemistry Department
P.O. Box 1892
Houston, TX 77251-1892

MAX SEEL
University Michigan Technological
College of Sciences & Arts
1400 Townsend Drive
Houghton, MI 49931-1295

ELIZABETH SEIBERLING
University of Florida
Department of Physics
528 NSC
Gainesville, FL 32611

HIDEO SEKINO
University of Florida
Quantum Theory Project
370 Williamson Hall
Gainesville, FL 32611

JORGE SEMINARIO
University of New Orleans
Chemistry Department
New Orleans, LA 70148

ISAIAH SHAVITT
Ohio State University
Department of Chemistry
120 W. 18th Avenue
Columbus, OH 43210-1173

GEORGE SHIELDS
Lake Forest College
Chemistry Department
555 N. Sheridan
Lake Forest, IL 60045

HARRIS SILVERSTONE
Johns Hopkins University
Department of Chemistry
3400 N. Charles Street
Baltimore, MD 21218-2685

CARLOS SIMMERLING
University of Illinois
Chemistry Department, M/C 111
845 W. Taylor
Chicago, IL 60607-7061

OKTAY SINANOGLU
Yale University
Chemistry Department, SCL
P.O. Box 6666
New Haven, CT 06511-8118

VEDENE SMITH, JR.
Queen's University
Department of Chemistry
Kingston, Ontario K7L 3N6
CANADA

RICHARD SQUIRE
Marshall University
Department of Chemistry
901 W. DuPont Avenue
Belle, WV 25015

CHRISTOPHER STANTON
University of Florida
Department of Physics
Gainesville, FL 32611

JOHN STANTON
University of Texas at Austin
Department of Chemistry and Biochemistry
Austin, TX 78712-1167

KRASSIMIR STAVREV
University of Florida
Quantum Theory Project
380 WM Hall
Gainesville, FL 32611

BORIS STEFANOV
Florida State University
Department of Chemistry
Tallahassee, FL 32306-3006

WALTER STEVENS
Standards & Technology National
Institute of CARB
9600 Gudelsky Drive
Rockville, MD 20850

MALCOLM STOCKS
Oak Ridge National Laboratory
Materials Science Section
Metals and Ceramics Division
P.O. Box 2008
Oak Ridge, TN 37831-6114

JOYCE STOUT
Indianapolis Indiana-Purdue University
Chemistry Department
402 N. Blackford Street
Indianapolis, IN 46202

ALAN STRATTON
CAChe Scientific
P.O. Box 349
Chilhowie, VA 24319

SPERANTA SVIRSCHEVSKY
Florida State University
Inst. of Molecular Biophysics
Tallahassee, FL 32306-3015

PETER SZALAY
University of Florida
Quantum Theory Project
362 Williamson Hall, P.O. Box 118435
Gainesville, FL 32611-8435

KRZYSZTOF SZALEWICZ
University of Delaware
Department of Physics
Newark, DE 19716

LASZLO SZENTPALY
University of the West Indies
Chemistry Department
Mona Campus
Kingston 7
JAMAICA

JAMES TALMAN
University of Western Ontario
Applied Matematics Department
London, Ontario N6A 5B7
CANADA

TOOMAS TAMM
University of Florida
Department of Chemistry
378 Williamson Hall
P.O. Box 118435
Gainesville, FL 32611-8435

UFUK TANERI
University of Waterloo Quantum Theory Group
Department of Chemistry
T.R.N.C.
Gazi Magosa, Mersin 10
TURKEY

JOACHIM THEILHABER
Thinking Machines Corporation
245 First Street
Cambridge, MA 02142

WALTER THIEL
Universitat Zurich
Organisch-Chemisches Institut
Winterthurerstr. 190
Zurich CH-8057
SWITZERLAND

COLIN THOMSON
University of St. Andrews
Department of Chemistry
North Haugh
St. Andrews KY16 9ST
SCOTLAND

ERIK THULSTRUP
Roskilde University
Institute of Life Sciences
P.O. Box 260
Roskilde DK-4000
DENMARK

IGOR TOPOL
Program Resources, Inc. National Cancer
   Institute
Frederick Cancer Research & Development
   Center,
P.O. Box B
Frederick, MD 21702

SAMUEL B. TRICKEY
University of Florida
Quantum Theory Project
364 Williamson Hall
Gainesville, FL 32611-2085

NENAD TRINAJSTIC
The Rugjer Boskovic Institute
Physical Chemistry Dept.
Bijenicka 54
Zagreb, HR 41001
CROATIA

CARL TRINDLE
University of Virginia
Department of Chemistry
NCB 242, UVA
Charlottesville, Virginia 22903

DAN URRY
Birmingham University of Alabama
Lab. of Molecular Biophysics
525 Volker Hall, VH 300
Birmingham, AL 35294-0019

CARMELA VALDEMORO
Investigaciones Cientificas Consejo Superior de
   Instituto de Ciencias Material
Serrano 123
Madrid 28006
SPAIN

PIET VAN DUIJNEN
University of Groningen
Chemistry Department
Nyenborgh 4
Groningen, AG 9747
THE NETHERLANDS

MATESH VARMA
Department of Energy
ER-73
Washington, DC 20585

DAN VASILESCU
Universite de Nice
Laboratoire de Biophysique
Parc Valrose
Nice, Cedex 2 06108
FRANCE

DANIEL VERCAUTEREN
de la Paix Universite Notre Dame
Lab. Physico Chimie Info.
Rue de Bruxelles, 61
Namur B-5000
BELGIUM

ROBERT VERGENZ
University of North Florida
Department of Natural Sciences
Jacksonville, FL 32225-2645

SHAWN WALLACE
University of North Florida
12426 Valpariss Tr.
Jacksonville, FL 32223

JIAN WANG
Dalhousie University
Department of Chemistry
Halifax, Nova Scotia B3H 4J3
CANADA

ARIEH WARSHEL
California University of Southern
Chemistry Department
Los Angeles, CA 90089

DAVID WATT
University of St. Andrews
Physics & Astronomy Department
Bute Medical Bldgs.
St. Andrews, Fife KY16 9TS
SCOTLAND

JOHN WATTS
University of Florida
Quantum Theory Project
367 WM Hall
Gainesville, FL 32611

BRIAN WEINER
Penn State University
Physics Department
College Place
Dubois, PA 15801

CARTER WHITE
Naval Research Laboratory
Supervisory Research Physicist
Head Theor. Chem. Section
Washington, DC 20375

CHRISTOPHER WHITE
Berkeley University of California
Department of Chemistry
Berkeley, CA 94720

ANGELA WILSON
University of Minnesota Department of
   Chemistry
139 Smith Hall
207 Pleasant St., S.E.
Minneapolis, MN 55455-0431

PETER WINKLER
University of Nevada
Department of Physics, 220
Reno, NV 89557-0058

BRIAN WLADKOWSKI
Biotechnology Center for Advanced Research
9600 Gudelsky Dr.
Rockville, MD 20850

PETER WOLYNES
University of Illinois
Department of Chemistry
505 South Mathews Avenue
Urbana, IL 61801

JIN ZHONG WU
University of Florida
Quantum Theory Project
368 Williamson Hall
Gainesville, FL 32611

SATOSHI YABUSHITA
Chiba University
Chemistry Department
1-33, Yayoi-cho
Inage-ku, Chiba 263
JAPAN

JIAN-MIN YUAN
Drexel University
Physics Department
32nd & Chestnut Street
Philadelphia, PA 19104

MICHAEL ZERNER
University of Florida
Quantum Theory Project
382 Williamson Hall
Gainesville, FL 32611

CHANG-GUO ZHAN
University Central China Normal
Department of Chemistry
Wuhan 430070
CHINA

HONGXING ZHANG
Jilin University Inst. of Theoretical Chemistry
Changchun 130023
CHINA

QINGSHENG ZHAO
University of North Carolina
Department of Chemistry
Chapel Hill, North Carolina 27599

XUEHE ZHENG
University of Florida
Dept of Chemistry
440 CRB
Gainesville, FL 32611

ZHONGXIANG ZHOU
Protein Design Labs, Inc.
2375 Garcia Avenue
Mountain View, CA 94043

JAMES ZIEGLER
IBM Research
P.O. Box 218
Yorktown Heights, NY 10598

YU ZUB
Ukrainian Academy of Science
Institute of Surface Chemistry
pr. Nauki, 31
Kiev 252022
UKRAINE

Author Index

Published Symposia of the
International Journal of Quantum Chemistry

1967   QUANTUM CHEMISTRY SYMPOSIUM NO. 1
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory)

1968   QUANTUM CHEMISTRY SYMPOSIUM NO. 2
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1969   QUANTUM CHEMISTRY SYMPOSIUM NO. 3 PART 1
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1970   QUANTUM CHEMISTRY SYMPOSIUM NO. 3 PART 2
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1971   QUANTUM CHEMISTRY SYMPOSIUM NO. 4
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1971   QUANTUM CHEMISTRY SYMPOSIUM NO. 5
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1972   QUANTUM CHEMISTRY SYMPOSIUM NO. 6
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1973   QUANTUM CHEMISTRY SYMPOSIUM NO. 7
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1974   QUANTUM CHEMISTRY SYMPOSIUM NO. 8
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Statistics)
QUANTUM BIOLOGY SYMPOSIUM NO. 1
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1975   QUANTUM CHEMISTRY SYMPOSIUM NO. 9
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Statistics)
QUANTUM BIOLOGY SYMPOSIUM NO. 2
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1976   QUANTUM CHEMISTRY SYMPOSIUM NO. 10
        (Proceedings of the International Symposium on Atomic, Molecular, and Solid-State
        Theory and Quantum Statistics)
        QUANTUM BIOLOGY SYMPOSIUM NO. 3
        (Proceedings of the International Symposium on Quantum Biology and Quantum Phar-
        macology)

1977   QUANTUM CHEMISTRY SYMPOSIUM NO. 11
        (Proceedings of the International Symposium on Atomic, Molecular, and Solid-State
        Theory, Collision Phenomena, and Computational Methods)
        QUANTUM BIOLOGY SYMPOSIUM NO. 4
        (Proceedings of the International Symposium on Quantum Biology and Quantum Phar-
        macology)

1978   QUANTUM CHEMISTRY SYMPOSIUM NO. 12
        (Proceedings of the International Symposium on Atomic, Molecular, and Solid-State
        Theory, Collision Phenomena, and Computational Methods)
        QUANTUM BIOLOGY SYMPOSIUM NO. 5
        (Proceedings of the International Symposium on Quantum Biology and Quantum Phar-
        macology)

1979   QUANTUM CHEMISTRY SYMPOSIUM NO. 13
        (Proceedings of the International Symposium on Atomic, Molecular, and Solid-State
        Theory, Collision Phenomena, Quantum Statistics, and Computational Methods)
        QUANTUM BIOLOGY SYMPOSIUM NO. 6
        (Proceedings of the International Symposium on Quantum Biology and Quantum Phar-
        macology)

1980   QUANTUM CHEMISTRY SYMPOSIUM NO. 14
        (Proceedings of the International Symposium on Atomic, Molecular, and Solid-State
        Theory, Collision Phenomena, Quantum Statistics, and Computational Methods)
        QUANTUM BIOLOGY SYMPOSIUM NO. 7
        (Proceedings of the International Symposium on Quantum Biology and Quantum Phar-
        macology)

1981   QUANTUM CHEMISTRY SYMPOSIUM NO. 15
        (Proceedings of the International Symposium on Atomic, Molecular, and Solid-State
        Theory, Collision Phenomena, and Computational Quantum Chemistry)
        QUANTUM BIOLOGY SYMPOSIUM NO. 8
        (Proceedings of the International Symposium on Quantum Biology and Quantum Phar-
        macology)

1982   QUANTUM CHEMISTRY SYMPOSIUM NO. 16
        (Proceedings of the International Symposium on Quantum Chemistry, Theory of Con-
        densed Matter, and Propagator Methods in the Quantum Theory of Matter)
        QUANTUM BIOLOGY SYMPOSIUM NO. 9
        (Proceedings of the International Symposium on Quantum Biology and Quantum Phar-
        macology)

1983   QUANTUM CHEMISTRY SYMPOSIUM NO. 17
        (Proceedings of the International Symposium on Atomic, Molecular, and Solid-State
        Theory, Collision Phenomena, and Computational Quantum Chemistry)

QUANTUM BIOLOGY SYMPOSIUM NO. 10
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1984   QUANTUM CHEMISTRY SYMPOSIUM NO. 18
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory, and Computational Quantum Chemistry)
QUANTUM BIOLOGY SYMPOSIUM NO. 11
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1985   QUANTUM CHEMISTRY SYMPOSIUM NO. 19
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory, Scattering Problems, Many Body Phenomena, and Computational Quantum Chemistry)
QUANTUM BIOLOGY SYMPOSIUM NO. 12
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1986   QUANTUM CHEMISTRY SYMPOSIUM NO. 20
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory, Scattering Problems, Many Body Phenomena, and Computational Quantum Chemistry)
QUANTUM BIOLOGY SYMPOSIUM NO. 13
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1987   QUANTUM CHEMISTRY SYMPOSIUM NO. 21
(Proceedings of the International Symposium on Quantum Chemistry, Solid-State Theory, and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 14
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1988   QUANTUM CHEMISTRY SYMPOSIUM NO. 22
(Proceedings of the International Symposium on Quantum Chemistry, Solid-State Theory, and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 15
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1989   QUANTUM CHEMISTRY SYMPOSIUM NO. 23
(Proceedings of the International Symposium on Quantum Chemistry, Solid-State Theory, and Molecular Dynamics)
QUANTUM BIOLOGY SYMPOSIUM NO. 16
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1990   QUANTUM CHEMISTRY SYMPOSIUM NO. 24
(Proceedings of the International Symposium on Quantum Chemistry, Solid State Physics, and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 17
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1991   QUANTUM CHEMISTRY SYMPOSIUM NO. 25
(Proceedings of the International Symposium on Quantum Chemistry, Solid State Physics, and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 18
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1992   QUANTUM CHEMISTRY SYMPOSIUM NO. 26
(Proceedings of the International Symposium on Atomic, Molecular, and Condensed Matter Theory and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 19
(Proceedings of the International Symposium on the Application of Fundamental Theory to Problems of Biology and Pharmacology)

1993   QUANTUM CHEMISTRY SYMPOSIUM NO. 27
(Proceedings of the International Symposium on Atomic, Molecular, and Condensed Matter Theory and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 20
(Proceedings of the International Symposium on the Application of Fundamental Theory to Problems of Biology and Pharmacology)

1994   QUANTUM CHEMISTRY SYMPOSIUM NO. 28
(Proceedings of the International Symposium on Atomic, Molecular, and Condensed Matter Theory and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 21
(Proceedings of the International Symposium on the Application of Fundamental Theory to Problems of Biology and Pharmacology)

All of the above symposia can be individually purchased from the Subscription Department, John Wiley & Sons.